# BIOCHEMICAL AND STRUCTURAL STUDIES OF THE TRANSCRIPTION ELONGATION FACTOR SPT6 AND ITS BINDING PARTNERS

by

Matthew A. Sdano

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

The University of Utah

August 2016

# The University of Utah Graduate School

## STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Matthew A. Sdano**

has been approved by the following supervisory committee members:

| | | |
|---|---|---|
| **Christopher P. Hill** | , Chair | **6/7/2016** Date Approved |
| **Timothy G. Formosa** | , Member | **6/7/2016** Date Approved |
| **Bradley R. Cairns** | , Member | Date Approved |
| **Michael S. Kay** | , Member | **6/7/2016** Date Approved |
| **David M. Belnap** | , Member | **6/7/2016** Date Approved |

and by **Wesley I. Sundquist** , Chair/Dean of

the Department/College/School of **Biochemistry**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Production of a functional protein requires coordination and regulation of many factors that control different cellular processes. One of these factors, Spt6, is a highly conserved nuclear protein that has roles in several facets of gene expression. Spt6 is best known for its ability to chaperone histones and modulate chromatin structure, but it also functions as a transcription elongation factor and in mRNA processing and export. Spt6 co-localizes with elongating RNA polymerase II (RNAPII) where it reassembles nucleosomes following RNAPII passage in order to repress aberrant transcription initiation. Furthermore, Spt6 directly stimulates RNAPII elongation rates and coordinates co-transcriptional mRNA processing. Despite the wealth of functional data that implicate Spt6 in these processes, little is known about the mechanistic basis for these activities. In order to gain mechanistic understanding of Spt6 activities, the work presented in this thesis focused on biochemical, structural, and functional characterization of Spt6 interactions with other proteins.

These studies reveal the true Spt6 binding site on RNAPII, identify Tom1 as a novel Spt6 binding partner, and demonstrate a specific interaction between Spt6 and histones H3-H4 that is competitive with DNA. The elucidation of the authentic Spt6 binding site on RNAPII has allowed us to develop tools to probe the mechanism of Spt6 recruitment to transcribed regions. Identification of Tom1

as a phosphorylated Spt6 binding partner provides a physical link to potential Spt6 functions such as mRNA export, cell cycle regulation, and regulation of histone levels. Characterization of the interaction with H3-H4 brings us closer to a mechanistic understanding of the histone chaperone activity of Spt6. Overall, the data presented in this work advance our knowledge of the interactions that regulate Spt6 function and will aid future studies to further dissect the mechanistic basis for Spt6 roles in gene expression.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my co-workers, friends, and family for making this thesis possible. The Hill lab has been a wonderful environment to develop as a scientist. Chris has been an exceptional mentor, creating an environment where excellent science and ethical behavior thrive. I am grateful for the opportunity he provided me to develop as an independent scientist. My labmates, past and present, have been instrumental for my education. I am thankful for the numerous conversations discussing experimental details, recent seminars, and life. These conversations made the lab an enjoyable place to work. I would also like to thank members of my thesis committee, Tim Formosa, Brad Cairns, Michael Kay, and David Belnap, for the outstanding guidance they provided throughout my graduate school career. I would like to extend an extra thank you to Seth McDonald, Frank Whitby, David Belnap, Peter Shen, and Tim Formosa for their contributions to my project. Seth trained me in the lab and provided guidance in the initial stages of my project. Frank helped me with all things crystallographic. David and Peter trained me in cryo-EM and helped collect and process data. Tim has been like a second mentor, bringing a biological aspect to my project and providing experimental guidance. I am grateful for Tim's breadth of knowledge about biology and his willingness to answer questions at anytime. I would also like to thank Erik Kish-Trier for the hundreds of miles he ran with me

during his time in lab. I would also like to thank my family for their endless love and support. My parents' have always encouraged me to work hard, follow my passion, do the right thing, and live life to the fullest. Most importantly, I would like to thank my wife Mallory. I could not have asked for a more understanding person to be by my side during graduate school. From bringing me dinner when experiments ran late to hiking the Grand Canyon when I needed to unwind, she has been my encouragement and motivation. She is the reason I think fondly of my time as a graduate student. I would also like to acknowledge Elsevier Science for permission to use published articles and figures within this dissertation.

# CHAPTER 1

# INTRODUCTION

## Gene Expression

The central dogma of biology distills gene expression into its two most basic steps: transcription and translation. Transcription is the process by which the information encoded in a gene is read and copied into a messenger RNA (mRNA). Translation is the process by which the information in an mRNA is used to synthesize a functional protein from amino acids. In reality, gene expression is a complex, multistep process with multiple layers of regulation. Consequently, many factors exist to facilitate, coordinate, and regulate the many facets of gene expression. One of these factors, Spt6, functions in several key regulatory processes including modulation of chromatin structure and DNA accessibility, transcription initiation and elongation, and mRNA processing and export. This thesis focuses on Spt6 interactions and how they contribute to different facets of gene expression.

## Chromatin Structure and Dynamics

Every cell in the human body contains nearly 2 meters of DNA that are condensed into a nucleus with a diameter of about 10 μM (Marino-Ramirez et al., 2005). This remarkable packaging feat is accomplished by compacting DNA into

a nucleoprotein complex called chromatin. The most basic unit of chromatin is the nucleosome, which consists of DNA wrapped around core histone proteins (Luger et al., 1997). Nucleosomes facilitate assembly of higher order chromatin structures that have been divided into three hierarchical tiers of compaction (Figure 1-1) (Woodcock and Dimitrov, 2001). The first tier consists of nucleosomal arrays that are ~11 nm wide and resemble 'beads on a string.' Nucleosomal arrays are further compacted, with the aid of additional proteins such as histone H1, into the next tier of chromatin structure known as the 30 nm fiber (Thomas, 1999). The structure of the 30 nm fiber has been controversial (Dorigo et al., 2004; Finch and Klug, 1976; Robinson et al., 2006; Williams et al., 1986), but a recent cryo-EM reconstruction provides strong evidence supporting a two-start helix (Song et al., 2014). The third tier of chromatin structure consists of 300 and 700 nm fibers that are formed by inter-fiber interactions (Kan et al., 2007; Li and Reinberg, 2011).

While assembling DNA into chromatin is an effective packaging strategy, it also restricts access for DNA-templated processes including replication, repair, and transcription. As a result, chromatin state is intimately tied to biological processes. Therefore, packaging and unpacking of each tier of chromatin structure is dynamic and highly regulated to allow precise spatial and temporal access to DNA (Luger and Hansen, 2005). The dynamic assembly and disassembly of chromatin is made possible by the unique structure of the nucleosome.

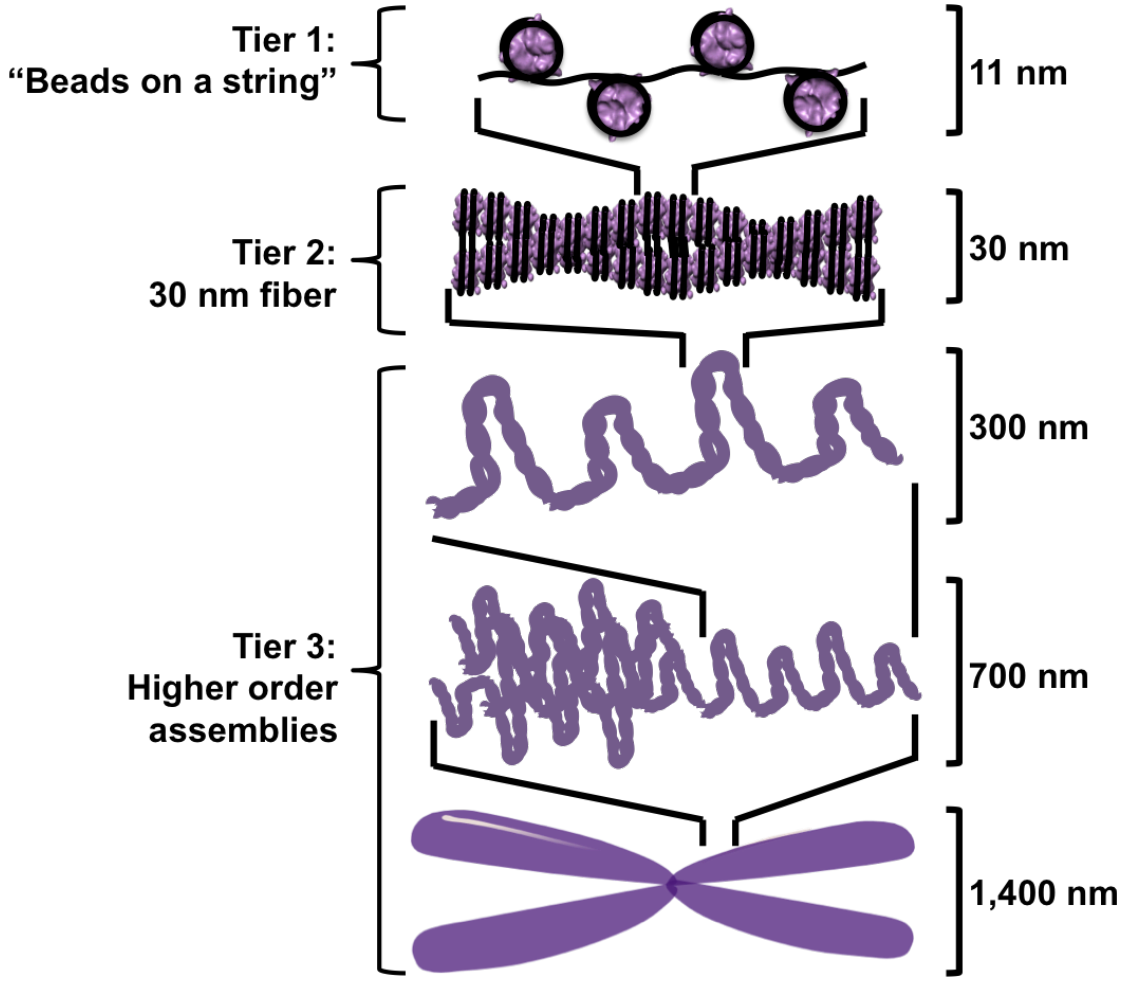



**Figure 1.1. Organization of the genome into chromatin.**
Schematic representation of the three tiers of chromatin compaction.

Nucleosome Structure

The composition and general shape of the nucleosome has been known for decades (Kornberg, 1977), but high-resolution structures remained elusive until the mid-1990s. In 1997, a 2.8 angstrom structure revealed the first atomic view of the nucleosome core particle (NCP) (Luger et al., 1997). The NCP consists of a histone octamer composed of two copies of each of the four core histones H2A, H2B, H3, and H4 that are wrapped by ~147 base pairs of DNA in a left-handed superhelix. Within the octamer, H2A dimerizes with H2B, and H3 dimerizes with H4, for a total of four histone pairs (Figure 1.2). The H3-H4 dimers further assemble into an H3-H4 heterotetramer [$(H3\text{-}H4)_2$] that is flanked by the H2A-H2B dimers to create a left-handed superhelix that templates the DNA.

The four core histones share a structurally similar helix-turn-helix-turn-helix motif that has been named the histone fold (Arents et al., 1991; Luger et al., 1997). The histone fold facilitates heterodimeric interactions between dimer pairs in what is known as the “handshake motif.” The $(H3\text{-}H4)_2$ tetramer is formed through a four-helix bundle between H3 and H3’, and the H2A-H2B dimers interact with $(H3\text{-}H4)_2$ through analogous 4-helix bundles between H2B and H4. In total, the histone octamer contacts 121 base pairs of DNA primarily through the phosphodiester backbone. The extensive protein-protein and protein-DNA interactions explain the stable structure of the nucleosome.

Disordered tails extend from the ordered histone folds and do not contribute to the structure of the NCP. The tails are rich in lysine, arginine, and serine residues that are posttranslationally modified with methyl, acetyl,

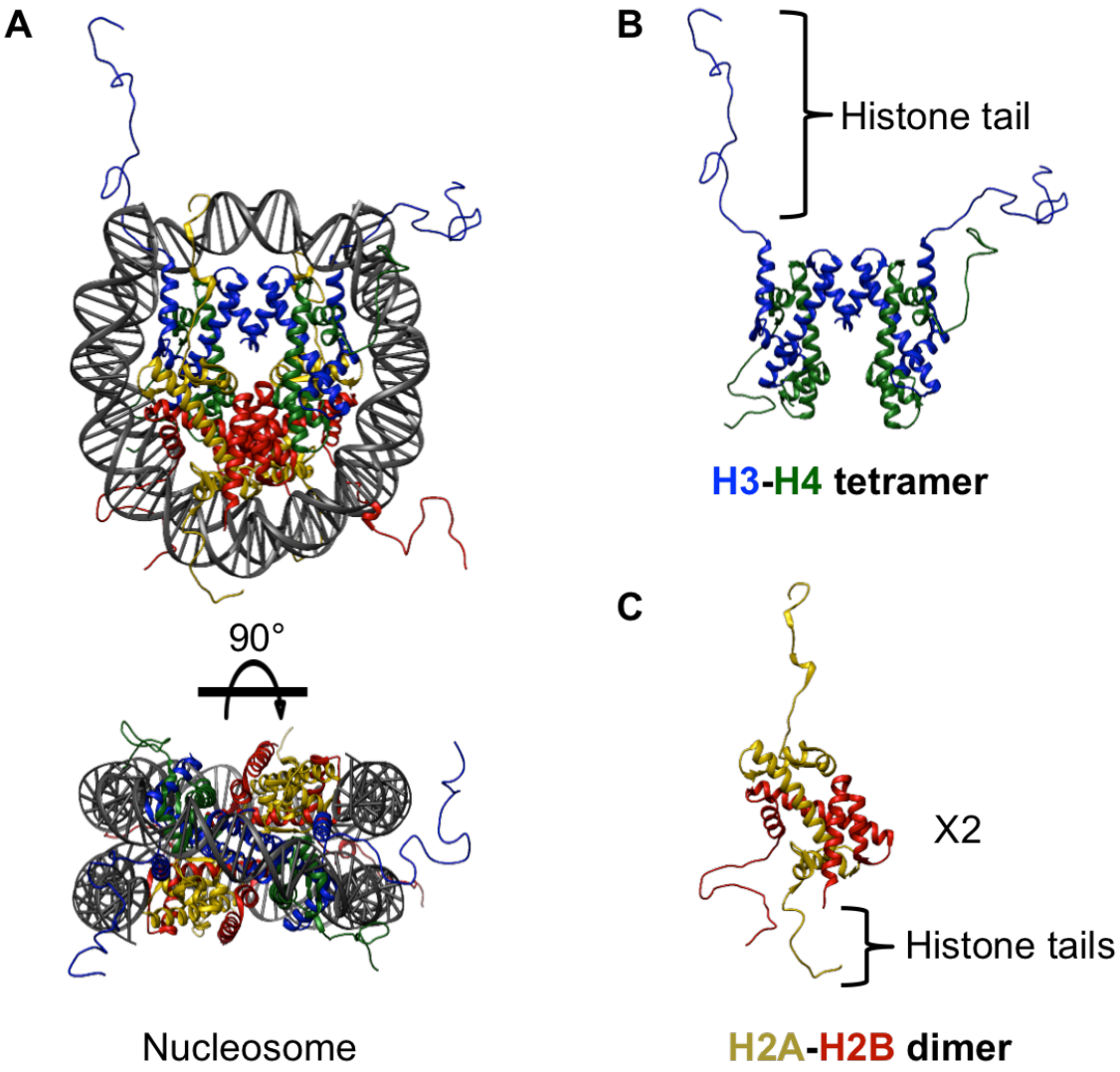



**Figure 1.2: Structure of the nucleosome.**
A) Orthogonal views of the nucleosome core particle represented in ribbons. H2A: yellow; H2B: red; H3: blue; H4: green; DNA: grey.
B and C) View of the H3-H4 tetramer and H2A-H2B dimer removed from the nucleosome. Histone tails are indicated.
Figure created using UCSF Chimera (Pettersen et al., 2004) using PDB ID: 1kx5 (Davey et al., 2002).

phosphate, ubiquitin, and other moieties to regulate chromatin compaction and recruitment of specific factors (Bannister and Kouzarides, 2011; Rothbart and Strahl, 2014). The histone tails and their modification states contribute to the assembly of higher order chromatin structures, with the H4 tail having the most pronounced effect (McBryant et al., 2009; Shogren-Knaak et al., 2006; Wang and Hayes, 2008). Generally speaking, histone posttranslational modifications (PTMs) regulate chromatin state either by recruiting effector proteins or by directly affecting nucleosome and chromatin structure (Li and Reinberg, 2011; Taverna et al., 2007). Effector proteins are recruited to chromatin through reader domains that recognize specific histone modifications in order to facilitate the coordinated recruitment of machinery that acts on chromatin. Modifications that alter chromatin structure directly tend to weaken inter- and intra-nucleosomal contacts by neutralizing positive charges on the highly basic histones (Shogren-Knaak et al., 2006). Because histone modifications have profound effects on chromatin structure, the histone modification state is intimately tied to regulation of DNA-templated processes.

In some cases, the canonical core histones are replaced by histone variants that resemble their canonical counterparts but differ in primary sequence. Sequence variations range from several amino acids to more dramatic changes such as the introduction of new domains (Mattiroli et al., 2015; Maze et al., 2014). H2A variants are the most diverse family of core histones while H2B and H3 variants are more similar to their counterparts. Variant histones alter the structural properties of the nucleosome to influence chromatin structure and

dynamics.

Like higher order chromatin structures, nucleosomes pose barriers to DNA-templated processes. Consequently, nucleosome position and stability are linked to biological function. Transcription is one of several processes that depends on the coordinated assembly, disassembly, and repositioning of nucleosomes.

## Transcription

Transcription is the process by which genetic information stored as DNA is copied as a molecule of RNA. In eukaryotes, the three enzyme complexes that catalyze RNA synthesis are called RNA polymerase (RNAP) I, II, and III. Each polymerase transcribes a different class of genes. RNAPI is responsible for transcription of ribosomal RNAs (rRNA), which are used in ribosomal assembly and account for approximately 50% of cellular transcription. RNAPII transcribes protein-coding genes to produce mRNA as well as DNA encoding small nuclear RNA (snRNA) and micro RNA (miRNA). mRNA is used as the template for protein synthesis whereas snRNA and miRNA have roles in splicing and post-transcriptional repression, respectively. RNAPIII transcribes small, structured RNAs such as 5S rRNA, tRNA, and spliceosomal U6 snRNAs that are important for ribosome biogenesis, protein synthesis, and splicing, respectively. Of the three polymerases, RNAPII is the best studied and most regulated.

RNAPII Structure and the CTD

RNAPII is a 12 subunit complex consisting of proteins Rpb1-Rpb12 (Figure 1.3A) (Cramer et al., 2000). Ten of the subunits form the catalytic core while the Rpb4 and Rpb7 subunits form a heterodimer that can dissociate from the enzyme (Edwards et al., 1991; Mosley et al., 2013). The largest subunit, Rpb1, contains an unstructured C-terminal domain (CTD) that consists of heptapeptide repeats of the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$. The CTD is subject to multiple reversible modifications including peptidyl-prolyl isomerization, glycosylation, and phosphorylation that create a dynamic binding surface that recruits chromatin regulators and mRNA processing factors. The modification state of the CTD is coupled to the transcription cycle, facilitating co-transcriptional coordination of chromatin state and mRNA processing (Figure 1.3B).

The best characterized of the CTD modifications are tyrosine, serine, and threonine phosphorylation. Recent mass spectrometry experiments determined that $S_2$ and $S_5$ phosphorylation are the most abundant modifications, consistent with their importance in coordinating the stages of the transcription cycle as well as co-transcriptional mRNA processing (Schuller et al., 2016; Suh et al., 2016). Chromatin immunoprecipitation (ChIP) experiments using antibodies that recognize specific phosphorylation states of the CTD revealed that $S_5$ phosphorylation is most abundant in the 5' region of genes whereas $S_2$ phosphorylation is most abundant in the 3' region (Hintermair et al., 2012; Mayer et al., 2012; Mayer et al., 2010). This is consistent with their roles in recruiting the

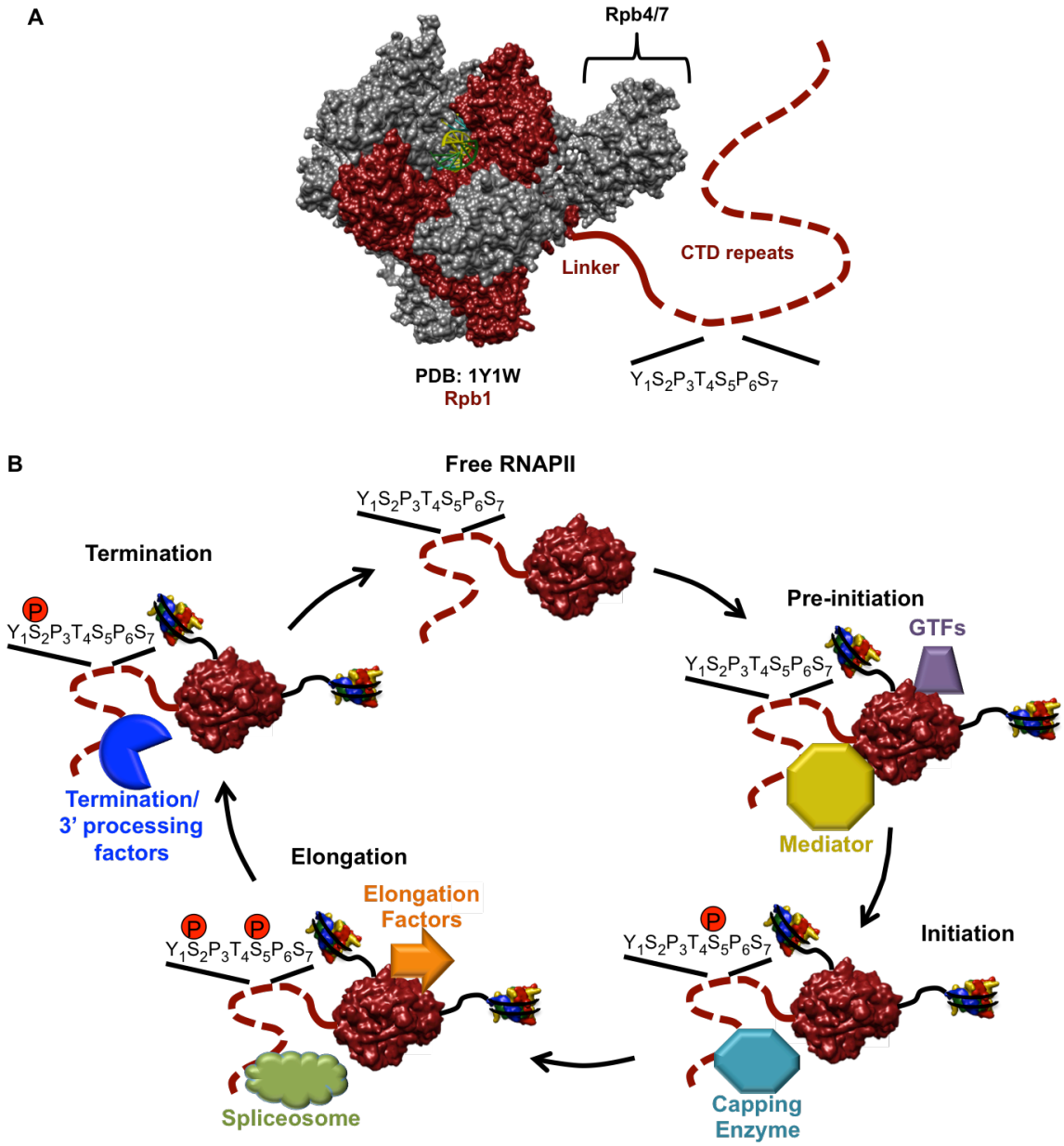



**Figure 1.3: Transcription by RNA polymerase II.**
A) Model of 12 subunit RNAPII with the Rpb1 linker and CTD. The core enzyme (PDB ID: 1Y1W) (Kettenberger et al., 2004) is shown as a surface representation with Rpb1 colored red and DNA colored yellow and cyan. The Rpb4/7 heterodimer is highlighted. The Rpb1 linker and CTD are represented by solid and dashed lines, respectively. Figure created using UCSF Chimera (Pettersen et al., 2004).
B) Overview of $S_2$ and $S_5$ phosphorylation of RNAPII CTD during the transcription cycle highlighting coordination of mRNA processing.

mRNA capping enzyme and termination factors, respectively.

The Rpb1 CTD is tethered to the RNAPII core through a flexible ~80 residue linker. The first 10 residues of the linker contact the Rpb4/7 heterodimer and are ordered in the *Saccharomyces cerevisiae* RNAPII crystal structures (Armache et al., 2005). In *Schizosaccharomyces pombe*, an additional 47 residues of the linker are ordered due to additional contacts with Rpb4/7 (Spahr et al., 2009). Consequently, it has been suggested that the linker may influence transcription by stabilizing Rpb4/7 association with the core. While the CTD can support viability when fused to other RNAPII subunits, the linker is only functional when contiguous with Rpb1 (Suh et al., 2013). In Chapter 3 of this dissertation, I present evidence that a function of the Rpb1 linker is to bind Spt6.

Initiation

Transcription by RNAPII occurs in three phases: initiation, elongation, and termination. During the initiation phase, a pre-initiation complex (PIC) is assembled from the general transcription factors (GTFs) and RNAPII (Kornberg, 2007). The GTFs include TFIIB, -D, -E, -F, and –H and are necessary for promoter recognition and initiation of transcription. TFIIH contains multiple subunits with catalytic activities including a kinase module that phosphorylates $S_5$ of the RNAPII CTD (Akhtar et al., 2009). The exact architecture of the PIC is controversial, as two conflicting cryo-EM structures have been published from different groups (He et al., 2013; Murakami et al., 2013).

RNAPII that enters the PIC has an unphosphorylated CTD, which is a

binding surface for the multiprotein Mediator complex (Myers et al., 1998). Mediator stabilizes the PIC, communicates signals from transcription factors to RNAPII, and stimulates kinase activity of TFIIH (Sogaard and Svejstrup, 2007). Phosphorylation of $S_5$ of the CTD by TFIIH recruits the mRNA capping enzyme and correlates with the transition from initiation to elongation (Ho and Shuman, 1999).

Elongation

Transcription elongation is the processive addition of nucleotides to the nascent RNA chain. The elongation phase is characterized by a hyper-phosphorylated CTD and association of transcription elongation and mRNA processing factors. The mRNA capping enzyme is recruited to RNAPII through direct interactions with the $S_5$ phosphorylated CTD (Ho and Shuman, 1999) and catalyzes addition of a 5' cap shortly after the nascent mRNA reaches the RNAPII surface (Rasmussen and Lis, 1993). The 5' cap consists of a guanine nucleotide connected to the 5' end of a transcript through a 5' to 5' triphosphate bond. The cap stabilizes the mRNA and protects against exonuclease digestion (Shuman, 2001).

Following mRNA capping, the CTD phosphorylation state shifts from $S_5$ phosphorylation to $S_2$ phosphorylation (Mayer et al., 2010). This transition is mediated by activities of the Ssu72 phosphatase that dephosphorylates $S_5$ (Krishnamurthy et al., 2004) and the cyclin-dependent kinases Ctk1 and Bur1 that phosphorylate $S_2$ (Bowman and Kelly, 2014). Transcription elongation factors are recruited through direct interactions with the RNAPII core enzyme and

the $S_2$ phosphorylated CTD. Elongation factors including Spt4/5, Polymerase Associated Factor 1 complex (Paf1C), FACT, and Spt6 enhance processivity by directly acting on RNAPII and/or influencing chromatin structure to allow RNAPII passage (Endoh et al., 2004; Hartzog et al., 1998; Kim et al., 2010; Orphanides et al., 1998).

During the elongation phase, a large ribonucleoprotein complex called the spliceosome catalyzes co-transcriptional splicing of the nascent RNA (Ardehali and Lis, 2009; Brugiolo et al., 2013; Osheim et al., 1985). Splicing is the removal of introns from the primary transcript, leaving only the protein coding sequence. The phosphorylated RNAPII CTD interacts directly with several splicing factors and is necessary for spliceosome assembly and proper splicing (Emili et al., 2002; Hirose et al., 1999; Morris and Greenleaf, 2000). Therefore, CTD phosphorylation coordinates the stages of the transcription cycle with the splicing machinery.

Termination

The end of transcription elongation is signaled by passage through the poly(A) site, which triggers transcript cleavage and is a prerequisite for termination (Connelly and Manley, 1988; Logan et al., 1987). Termination occurs anywhere from several bases to several kilobases downstream from the poly(A) site (Hagenbuchle et al., 1984). Coincident with termination, the CTD is dephosphorylated, allowing recycling of RNAPII for subsequent rounds of initiation and elongation (Cho et al., 1999; Lian et al., 2008).

Two models, the allosteric model and the torpedo model, have been

proposed to describe how RNAPII is released from DNA. According to the allosteric model, transcription of the poly (A) signal causes conformational changes in the polymerase that destabilize the interaction with DNA, leading to decreased processivity followed by disassembly (Logan et al., 1987). In the torpedo model, the 5' phosphate on the cleaved RNA recruits an exonuclease that degrades the RNA and stimulates DNA release upon catching up to the polymerase (Connelly and Manley, 1988). Because both models are supported by data, the current view is that the allosteric and the torpedo mechanisms both contribute to transcription termination (Richard and Manley, 2009).

Transcription of the poly(A) site also signals cleavage and polyadenylation of the transcript. Polyadenylation is the addition of multiple adenosine monophosphates to the 3' end of an mRNA and is important for nuclear export, translation, and stability of the mRNA (Colgan and Manley, 1997). Like mRNA capping and splicing, cleavage and polyadenylation are coupled to the transcription cycle via the RNAPII CTD (de Almeida and Carmo-Fonseca, 2008). The $S_2$ phosphorylated CTD directly interacts with several components of the cleavage-polyadenylation machinery in order to coordinate recruitment with termination (Lunde et al., 2010; Meinhart and Cramer, 2004). This helps ensure that only completely synthesized mRNAs are processed and become export competent.

## Transcription Through Nucleosomes

Nucleosomes pose a physical barrier to transcription initiation and elongation. Assembly of the PIC requires chromatin to be in an open

conformation. Consequently, many promoters are void of nucleosomes to allow transcription factors access to DNA binding site (Bernstein et al., 2004; Lee et al., 2004; Sekinger et al., 2005). Alternatively, some genes, such as *PHO5* and *PHO8*, use strategically placed nucleosomes in the promoter to repress transcription until the nucleosomes are removed or repositioned (Adkins and Tyler, 2006). Nucleosomes in the coding region also impede transcription by preventing RNAPII passage during elongation (Lorch et al., 1992). Because nucleosomes have such profound effects on transcriptional output, they present an important mechanism for transcriptional regulation. Therefore, both initiation and elongation require remodeling, repositioning, or removal of nucleosomes to allow assembly of the PIC or passage of RNAPII. In addition, nucleosomes must be reassembled and repositioned following RNAPII passage in order to restore the default repressive chromatin state. Two classes of proteins, chromatin remodelers and histone chaperones, facilitate these activities.

Chromatin Remodelers

Chromatin remodelers hydrolyze ATP to slide, eject, assemble, or restructure nucleosomes to alter their position, occupancy, and composition. These activities control the packaging and unpackaging of chromatin to regulate DNA access for chromosomal processes. All remodelers share a similar ATPase domain, but unique flanking domains and associated subunits provide specialization (Clapier and Cairns, 2009). Remodelers are divided into four distinct families that are specialized for different functions and biological contexts. Domains that recognize specific histone modifications contribute to specialization

by linking remodeler activity to posttranslational modification states (Bannister and Kouzarides, 2011; Clapier and Cairns, 2009). The influence of remodelers on nucleosomes and chromatin not only influences transcription, but also has implications for chromatin assembly, DNA replication, and DNA repair and recombination (Clapier and Cairns, 2009).

## Histone Chaperones

Histone chaperones are a structurally diverse class of proteins that use binding energy to assemble or disassemble nucleosomes in order to regulate DNA accessibility during DNA-templated processes (De Koning et al., 2007; Laskey et al., 1978). Because histones are highly basic, non-nucleosomal histones are prone to aberrant interactions with DNA and acidic proteins. These interactions lead to aggregation and cell toxicity (Meeks-Wagner and Hartwell, 1986). To prevent nonproductive histone-DNA interactions, histone chaperones have highly acidic regions that interact with the basic histone surface to neutralize the charges. The coordinated activities of multiple histone chaperones facilitates the stepwise assembly and disassembly of nucleosomes (Das et al., 2010).

Histone chaperones are often classified by their specificity for different histones (Gurard-Levin et al., 2014). Typically, chaperones bind either H2A-H2B or H3-H4, although Spt6 and FACT have been shown to bind both (Bortvin and Winston, 1996; Formosa, 2012; Kemble et al., 2013; McCullough et al., 2015). In addition, some chaperones have specificity for histones variants (Mattiroli et al., 2015). Histone chaperones use a variety of structural motifs to interact with

histones, but structural studies have revealed common themes (Elsasser, 2013; Hondele and Ladurner, 2011; Kemble et al., 2015). Notably, chaperones tend to bind histone surfaces that are buried in a fully assembled nucleosome. Therefore, it appears that histone chaperones ensure ordered nucleosome assembly and disassembly by competing with histone-DNA and histone-histone contacts to stabilize assembly intermediates (Das et al., 2010; Hondele and Ladurner, 2011).

Chaperones function in many biological processes including DNA replication, transcription, repair, and heterochromatic silencing (Gurard-Levin et al., 2014). During replication and transcription, nucleosomes must be disassembled in front of a polymerase and reassembled behind. The similar topology of these processes allows some chaperones to function in both replication and transcription. Other chaperones, such as Spt6, are specific to transcription (Duina, 2011).

## Spt6

Spt6 was initially identified by Fred Winston in a genetic screen for suppressors of a Ty element insertion at the 5' end of the *HIS4* gene (Winston et al., 1984). Ty insertion at the *HIS4* gene causes transcriptional repression and the inability to grow on media lacking histidine. Suppressors of Ty insertions relieve transcriptional repression of *HIS4*, allowing growth on media lacking histidine. This phenomenon is designated the $Spt^-$ phenotype, and is shared by all members of the Spt gene family. Since nucleosomes are essential for global transcriptional repression, the $Spt^-$ phenotype often indicates loss of chromatin

integrity (Hainer et al., 2011; Kaplan et al., 2003). Therefore, it is unsurprising that Spt6 functions in global transcriptional regulation (Clark-Adams and Winston, 1987; Denis and Malvar, 1990; Neigeborn et al., 1987; Swanson and Winston, 1992).

Since Spt6 was identified, it has been implicated in a variety of biological processes throughout eukaryotes. Spt6 is important for embryogenesis in Zebrafish (Keegan et al., 2002; Kok et al., 2007), multiple stages of development in *Drosophila* (Ardehali et al., 2009), gut morphogenesis in *C. elegans* (Nishiwaki et al., 1993), and HIV transcriptional regulation in mammals (Vanti et al., 2009; Yoh et al., 2007). The broad importance of Spt6 is a consequence of diverse roles in gene expression including chaperoning histones, transcription elongation, and mRNA processing and export (Figure 1.4).

Spt6 Is a Histone Chaperone

The Spt$^-$ phenotype caused by *spt6* mutants is attributed to the requirement of Spt6 for maintaining repressive chromatin. The first evidence that Spt6 regulates chromatin state came from genetics. Mutations in *SPT6* suppress mutations in the Swi/Snf chromatin remodeling complex similar to deletion of one copy of the *H2A-H2B* gene pairs (Bortvin and Winston, 1996; Hirschhorn et al., 1992; Swanson and Winston, 1992). Spt6 is now known to directly bind H2A-H2B and H3-H4 and to promote nucleosome assembly *in vitro* (Bortvin and Winston, 1996; McCullough et al., 2015). Consistent with this, *spt6* mutations alter chromatin structure in the regulatory region of several genes, leading to spurious activation of transcription (Adkins and Tyler, 2006; Bortvin and Winston, 1996;

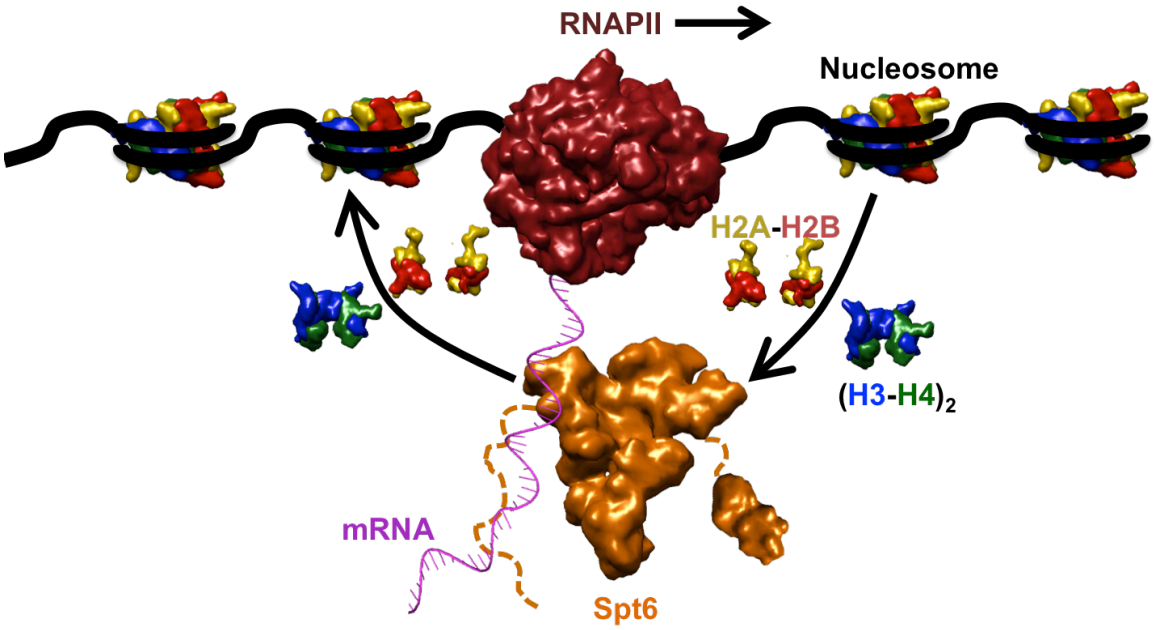



**Figure 1-4: Overview of Spt6 roles in gene expression.**
Spt6 is implicated in three distinct functions in gene expression. It is a transcription elongation factor that stimulates RNAPII elongation rates on naked DNA *in vitro*. It is also a histone chaperone that reassembles nucleosomes in the wake of elongating RNAII. It is also implicated in mRNA processing and export. Figure created using UCSF Chimera (Pettersen et al., 2004).

Hainer et al., 2011). The role of Spt6 in maintaining repressive chromatin extends beyond promoters to coding regions. Impaired Spt6 function results in decreased nucleosome occupancy throughout transcribed regions, leading to aberrant transcription initiation from cryptic promoters (Adkins and Tyler, 2006; DeGennaro et al., 2013; Ivanovska et al., 2011; Kaplan et al., 2003). Therefore, Spt6 appears to maintain repressive chromatin by reassembling nucleosomes in both promoters and in coding regions following RNAPII passage.

Spt6 Is a Transcription Elongation Factor

While the histone chaperone function of Spt6 undoubtedly impacts transcription elongation, Spt6 also directly affects the transcription machinery. *In vivo*, Spt6 stimulates elongation rates even when chromatin is in a permissive state (Ardehali et al., 2009). Furthermore, Spt6 enhances RNAPII elongation rates *in vitro* on DNA lacking nucleosomes (Endoh et al., 2004; Yoh et al., 2007). The role of Spt6 in transcription elongation is also supported by genetic evidence. *spt6* mutants display phenotypes associated with impaired transcription elongation such as sensitivity to 6-Azauracil (6-AU) and synthetic growth defects with transcription elongation factors (Hartzog et al., 1998; Sun et al., 2010; Winston et al., 1984). Consistent with roles for Spt6 in transcription elongation and co-transcriptional nucleosome assembly, Spt6 physically associates with elongating RNAPII and localizes to genes with a pattern similar to RNAPII (Andrulis et al., 2000; Kaplan et al., 2000; Kim et al., 2004; Mayer et al., 2010; Perales et al., 2013).

Spt6 Is Involved in mRNA Processing and Export

In addition to nucleosome assembly and transcription elongation, Spt6 also functions in mRNA processing and export. In yeast, *spt6* mutants are defective in transcription termination, 3' processing, and mRNA export (Bucheli and Buratowski, 2005; Estruch et al., 2009; Kaplan et al., 2005). Consistent with these observations, Spt6 genetically interacts with the mRNA export factors Mex67 and Yra1 (Burckin et al., 2005; Estruch et al., 2009). In mammals, Spt6 is required for nuclear export of mRNA as well as proper splicing of an HIV transcript (Yoh et al., 2007). Furthermore, Spt6 physically associates with the RNA processing exosome in *Drosophila* (Andrulis et al., 2002). It is unclear how Spt6 functions in mRNA processing and export, but it likely facilitates co-transcriptional recruitment of mRNA processing and export factors such as Spn1 and Yra1 (Yoh et al., 2007).

Spt6 Structure

Spt6 comprises 1,451 residues that can be divided into three structural regions: a disordered N-terminal region, a structured core, and a C-terminal tandem Src homology 2 (tSH2) domain (Close et al., 2011). The disordered N-terminal region comprises ~300 residues and is highly acidic, which is a common feature of histone chaperones. Residues 239-268 in this region interact with Spn1 and are necessary for binding nucleosomes (Diebold et al., 2010a; McDonald et al., 2010). Interestingly, these interactions are competitive, suggesting a switch mechanism in which Spn1 regulates Spt6 interaction with nucleosomes (McDonald et al., 2010). Consistent with this model, mutations in

histones H2A-H2B suppress temperature sensitivity caused by mutations in the Spt6-Spn1 interface (McCullough et al., 2015). While it is clear that the Spt6 N-terminus is required for chaperoning histones, the mechanistic basis of this activity is not understood. In order to understand how Spt6 functions as a chaperone, it is necessary to characterize how Spt6 interacts with the different components of the nucleosome. In Chapter 5 of this dissertation, I present evidence that the Spt6-Spn1 binding region is sufficient for binding histones H3-H4 and that this interaction is competitive with DNA.

The core of Spt6 contains multiple domains that are structurally similar to protein binding and DNA binding domains and has overall similarity to the prokaryotic transcription factor *Tex* (Close et al., 2011; Johnson et al., 2008). The Spt6 core structure is discussed in depth in Chapter 2. Briefly, recognizable structural motifs in the Spt6 core include a helix-turn-helix domain, a catalytically inactive YqgF domain, a helix-hairpin-helix domain, a death-like domain, and an S1 domain. The Spt6 core is capable of binding DNA *in vitro*; however, it is unclear if this interaction takes place *in vivo* (Close et al., 2011). Despite structural knowledge of the Spt6 core, the biological function is unclear. Structural similarity to *Tex* suggests an activity that is conserved among prokaryotes and eukaryotes, such as transcription on nucleosome-free DNA.

The C-terminus of Spt6 adopts a novel tSH2 domain that comprises the only recognizable SH2 domains in yeast (Close et al., 2011; Diebold et al., 2010b; Sun et al., 2010). Spt6 tSH2 consists of two SH2 domains stacked rigidly on top of one another. The N-terminal SH2 domain (nSH2) retains the primary

determinants of phosphotyrosine binding, including an invariant arginine (R1282) that is essential for phosphotyrosine binding in canonical SH2 domains. The C-terminal SH2 domain (cSH2), on the other hand, appears to be cryptic because residues critical for phosphotyrosine binding in canonical SH2 domains are substituted by amino acids with different chemical properties. The Spt6 tSH2 domain is important for Spt6 function and is the main focus of Chapters 3 and 4 of this dissertation.

## Spt6 tSH2 Domain

The tSH2 domain functions in most Spt6 activities including repressive chromatin maintenance, transcription elongation and termination, and mRNA processing and export. Deletion of the tSH2 domain in yeast causes slow growth, the $Spt^-$ phenotype, and phenotypes attributed to defects in transcription elongation (Diebold et al., 2010b; Hartzog et al., 1998; McCullough et al., 2015; Sun et al., 2010). In mammals, the tSH2 domain is implicated in class switch recombination, mRNA processing and export, and transcription elongation (Begum et al., 2012; Endoh et al., 2004; Okazaki et al., 2011; Yoh et al., 2007). Furthermore, the tSH2 domain is required for maximum Spt6 occupancy throughout transcribed genes (Burugula et al., 2014; Mayer et al., 2010).

The leading model for Spt6 tSH2 function is that it recruits Spt6 to RNAPII through direct interactions with the hyperphosphorylated CTD. This model originated from work in the Jones laboratory that showed direct binding between Spt6 tSH2 and purified, hyperphosphorylated RNAPII (Yoh et al., 2007). The interaction was recapitulated by purified GST-CTD that was phosphorylated *in*

*vitro* by the CTD kinase P-TEFb (Yoh et al., 2008). Prompted by these observations, we used fluorescence anisotropy to quantify binding between Spt6 tSH2 and CTD-derived peptides with different phosphorylation states (Close et al., 2011). The measured dissociation constants ($K_D$) for peptides with a single phosphorylated residue per repeat ranged from 110 to 250 μM. A peptide that was phosphorylated on both $S_2$ and $S_5$ of the CTD repeats had an affinity of 23 μM, but this peptide may be more prone to nonspecific interactions due to the additional negative charges from the extra phosphates. Regardless, these affinities are much weaker than expected for SH2 domains binding to phosphorylated peptides, which typically bind with a $K_D$ between 0.2 and 1 μM (Machida and Mayer, 2005). In addition, mutating Spt6 residues that are critical for phosphotyrosine binding in canonical SH2 domains decreased affinity by only 4 fold. These observations are consistent with data from other groups who used fluorescence anisotropy, pull-downs, and NMR to characterize Spt6 tSH2 binding to phosphorylated CTD-derived peptides (Diebold et al., 2010b; Liu et al., 2011; Mayer et al., 2012; Sun et al., 2010). One group did observe affinities of ~1 μM using fluorescence anisotropy, but the tighter affinity is likely due to abnormally low salt concentrations (10 mM NaCl) used in their assays (Mayer et al., 2012; Sun et al., 2010). Interestingly two of the studies raised the possibility of an additional phosphate-binding pocket on the back side of the non-canonical cSH2 domain. This pocket is lined with positively charged residues including a lysine that has been conserved throughout evolution. The residue displays a shift in an NMR chemical shift perturbation experiment (Liu et al., 2011), and mutating it

caused a modest reduction in affinity (~2.4 fold) for a CTD-derived phosphorylated peptide (Sun et al., 2010). Collectively, the data from these studies hint at the presence of a non-canonical phosphate-binding pocket and suggest that the phosphorylated CTD may not be the biological binding partner of Spt6 tSH2. However, the field has largely accepted the model that Spt6 tSH2 binds the RNAPII CTD. In Chapters 3 and 4 of this dissertation, I present evidence that the authentic Spt6 tSH2 binding partners are the RNAPII linker and the E3 ubiquitin ligase Tom1.

## Tom1

Tom1 is a 375 kDa E3 ubiquitin ligase with a C-terminal HECT domain and is homologous to mammalian Huwe1 (also called Mule, Lasu, Ureb1, Arf-BP1, HectH9, $E3^{Histone}$, Ptr1, Upl1, and Eel1). In yeast, deletion of Tom1 causes severe temperature sensitivity, nuclear accumulation of mRNA, and defects in nuclear division (Utsugi et al., 1999; Utsugi et al., 1995). In humans, mutations in the gene encoding Huwe1 are associated with cancer and intellectual disability (Froyen et al., 2012; Froyen et al., 2008; Inoue et al., 2013; Isrie et al., 2013; Liu et al., 2012; Nava et al., 2012; Zhao et al., 2009). The importance of Tom1 can be attributed to its roles in mRNA export, degradation of excess histones, transcription, and cell cycle regulation (Duncan et al., 2000; Iglesias et al., 2010; Kim and Koepp, 2012; Kim et al., 2012; Liu et al., 2005; Saleh et al., 1998; Singh et al., 2009; Utsugi et al., 1999). The role of Tom1 in mRNA export has been ascribed to ubiquitylation of the mRNA export adaptor Yra1 in order to facilitate its removal from mRNA ribonucleoprotein (mRNP) complexes prior to nuclear

export (Iglesias et al., 2010). Tom1 has also been reported to ubiquitylate histones, marking them for proteasomal degradation (Liu et al., 2005; Singh et al., 2009). As accumulation of excess non-nucleosomal histones is toxic to the cell, degradation via the ubiquitin proteasome system is an important mechanism for regulating free histone levels (Gunjan and Verreault, 2003; Meeks-Wagner and Hartwell, 1986). It is unclear how Tom1 functions in transcription, but it may involve ubiquitylation of components of the transcription coactivator complex SAGA (Saleh et al., 1998). Tom1 also controls degradation of Dia2 and Cdc6 during the G1 phase of the cell cycle, linking Tom1 to cell cycle regulation (Kim and Koepp, 2012; Kim et al., 2012). Despite some functional overlap between Spt6 and Tom1, it is unclear how a direct interaction between the proteins contributes to cellular physiology. The biochemistry presented in Chapter 4 has produced tools that can now be used to dissect the biological importance of this interaction.

## Goals of This Dissertation

Spt6 is conserved throughout eukaryotes, is essential for viability in *S. cerevisiae*, and has several roles in gene expression including transcription elongation, chaperoning histones, and mRNA processing and export. The importance of Spt6 in these processes is well documented, but the mechanistic details are unclear. The goal of this dissertation is to characterize Spt6 interactions with other proteins in order to understand how they relate to Spt6 function. The primary objectives were to identify biologically relevant binding partners for the Spt6 tSH2 domain and characterize the interaction between Spt6

and histones. The biochemical and structural data presented demonstrate that Spt6 tSH2 directly binds RNAPII and Tom1 and that the Spt6-Spn1 binding region interacts with histones H3-H4. The data and assays presented lay the foundation for future studies to probe the regulation and biological significance of these interactions.

## Outline of Chapters

### Chapter 2: Crystal Structure of the *S. cerevisiae* Spt6 Core and C-terminal Tandem SH2 Domain

The work discussed in Chapter 2 was originally published in the May 13, 2011 issue of Journal of Molecular Biology and is reprinted here in the published format. The focus of this work was to characterize the Spt6 structure, DNA binding, and phosphorylated RNAPII CTD binding. Sean Johnson and Devin Close solved the structure of the Spt6 core, and Devin solved the structure of Spt6 tSH2. Devin and Seth McDonald performed the DNA binding assays. I performed the bulk of the fluorescence anisotropy assays characterizing tSH2 domain binding to phosphorylated CTD-derived peptides. These assays provided the motivation to look for other binding partners as described in Chapters 3 and 4.

### Chapter 3: Mechanism of Phosphorylation Dependent Binding Between Spt6 and RNAPII

Work in this chapter identifies the authentic Spt6 tSH2 binding site on RNAPII. We show that the Rpb1 linker, but not the CTD, is necessary for Spt6 tSH2 binding. Binding is enhanced by phosphorylation of a highly conserved

serine in the linker. A crystal structure of the Spt6 tSH2-Rpb1 linker complex reveals a novel phosphate-binding pocket that coordinates the phosphoserine. Mutating residues in the interface disrupted binding *in vitro* and caused defects in maintaining repressive chromatin *in vivo*. Overall, our results indicate that a direct interaction with the Rpb1 linker, but not the CTD, recruits Spt6 to sites of transcription where it functions to reassemble nucleosomes in the wake of RNAPII.

The work in this chapter was motivated by the binding studies discussed in Chapter 2. I performed all of the biochemical experiments presented. James Fulcher and Mark Petersen, in the laboratory of Michael Kay, synthesized the peptides. I solved the structure of the Spt6 tSH2-Rpb1 linker complex under the supervision of Frank Whitby. The laboratory of Tim Formosa performed all genetic experiments.

Chapter 4: Identification of Tom1 as a Novel Binding Partner of the Spt6 Tandem SH2 Domain

The work in this chapter was performed in parallel with the work in Chapter 3 to identify alternative binding partners for the Spt6 tSH2 domain. Motivated by the work in Chapter 1, I performed a pull-down followed by far western blot to identify direct Spt6 tSH2 binding partners. Mass spectrometry identified a co-purifying protein that also bound in a far western blot as the E3 ubiquitin ligase Tom1. Binding to Tom1 was found to be phosphorylation dependent and mediated through a central ~30 residue segment. Spt6 mutations that disrupt binding to RNAPII also disrupt binding to Tom1, and these

interactions are competitive. Deletion of the Tom1-Spt6 tSH2 binding region suppresses temperature sensitivity caused by Tom1 overexpression, suggesting a biological role for the interaction.

Since little is known about the Tom1 structure, we also determined an 8 Å reconstruction of full-length Tom1 using cryo-electron microscopy (cryo-EM). The structure reveals a lock washer shape composed primarily of helical repeats. Despite the paucity of recognizable domains, ~70-85% of the entire Tom1 sequence is resolved. This work indicates that Tom1 is amenable to structure determination by cryo-EM and motivates pursuit of additional structures of Tom1 complexes.

I performed all of the biochemical experiments presented. Mass spectrometry experiments were performed by the University of Utah core facility. James Fulcher and Mark Petersen synthesized the peptides. Cryo-EM data collection and analysis was performed with the assistance of Peter Shen and David Belnap. The laboratory of Tim Formosa created the Tom1 overexpression yeast strain and performed the genetic experiments.

Chapter 5: Biochemical and Functional Characterization of the Spt6 Interaction with Histones H3-H4

The focus of this chapter was characterization of the Spt6 interaction with histones H3-H4. We show that the same Spt6 region that binds Spn1 is sufficient but not necessary for binding histones H3-H4. We identified mutations that disrupt the interaction with H3-H4 but not Spn1 *in vitro*, and these mutations cause phenotypes *in vivo*. We also find that Spt6 competes with DNA for H3-H4

binding. Overall, these results suggest that the Spt6-Spn1 binding region plays an important role in binding histones, but other binding surfaces likely exist.

This work was performed in close collaboration with Seth McDonald, who initiated these studies. Seth developed the fluorescence anisotropy binding assays and performed the DNA competition, ITC, and GST pull-down experiments. I performed the direct fluorescence anisotropy binding assays and identified mutations in the Spt6-Spn1 binding region that disrupt binding to H3-H4. Members of the lab of Tim Formosa performed the genetic experiments.

Chapter 6: Conclusions and Ongoing Research

Chapter 6 summarizes the findings of Chapters 2-5 and discusses the direction of future research, which will focus on Spt6 interactions and Tom1 function. Further characterization of the phosphorylation events that regulate Spt6 interactions with RNAPII and Tom1 will help define how Spt6 is recruited to various processes. Towards this goal, we have raised an antibody that specifically recognizes the serine phosphorylated Rpb1 linker, and we will use it to determine the abundance of this modification throughout the transcription cycle. In addition, it will be of interest to identify the kinase(s) and phosphatase(s) that catalyze phosphorylation and dephosphorylation. I also show preliminary biochemical data suggesting that Spt6 can interact directly with the RNAPII core in the absence of the Rpb1 linker, potentially explaining the role of Spt6 in transcription elongation. I discuss efforts to characterize this interaction using cryo-EM. The unanticipated interaction between Spt6 and Tom1 raises questions about the functional importance of the interaction as well as how Tom1

participates in a variety of biological processes. I discuss experiments designed to dissect the biological relevance of the Spt6-Tom1 interaction. Finally, I suggest additional biochemical experiments to more thoroughly characterize Spt6 interactions with histone H3-H4 and H2A-H2B. Overall, the experiments described in this chapter will build on data in this dissertation to further understanding of the interactions that contribute to the diverse roles of Spt6 in gene expression.

## References

Adkins, M.W., and Tyler, J.K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. Mol Cell *21*, 405-416.

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Mol Cell *34*, 387-393.

Andrulis, E.D., Guzman, E., Doring, P., Werner, J., and Lis, J.T. (2000). High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. Genes Dev *14*, 2635-2649.

Andrulis, E.D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., and Lis, J.T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. Nature *420*, 837-841.

Ardehali, M.B., and Lis, J.T. (2009). Tracking rates of transcription and splicing in vivo. Nat Struct Mol Biol *16*, 1123-1124.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J *28*, 1067-1077.

Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci U S A *88*, 10148-10152.

Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. J Biol Chem *280*,

7131-7134.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res *21*, 381-395.

Begum, N.A., Stanlie, A., Nakata, M., Akiyama, H., and Honjo, T. (2012). The histone chaperone Spt6 is required for activation-induced cytidine deaminase target determination through H3K4me3 regulation. J Biol Chem *287*, 32415-32429.

Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. Genome Biol *5*, R62.

Bortvin, A., and Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. Science *272*, 1473-1476.

Bowman, E.A., and Kelly, W.G. (2014). RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases. Nucleus *5*, 224-236.

Brugiolo, M., Herzel, L., and Neugebauer, K.M. (2013). Counting on co-transcriptional splicing. F1000Prime Rep *5*, 9.

Bucheli, M.E., and Buratowski, S. (2005). Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. EMBO J *24*, 2150-2160.

Burckin, T., Nagel, R., Mandel-Gutfreund, Y., Shiue, L., Clark, T.A., Chong, J.L., Chang, T.H., Squazzo, S., Hartzog, G., and Ares, M., Jr. (2005). Exploring functional relationships between components of the gene expression machinery. Nat Struct Mol Biol *12*, 175-182.

Burugula, B.B., Jeronimo, C., Pathak, R., Jones, J.W., Robert, F., and Govind, C.K. (2014). Histone deacetylases and phosphorylated polymerase II C-terminal domain recruit Spt6 for cotranscriptional histone reassembly. Mol Cell Biol *34*, 4115-4129.

Cho, H., Kim, T.K., Mancebo, H., Lane, W.S., Flores, O., and Reinberg, D. (1999). A protein phosphatase functions to recycle RNA polymerase II. Genes Dev *13*, 1540-1552.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273-304.

Clark-Adams, C.D., and Winston, F. (1987). The SPT6 gene is essential for growth and is required for delta-mediated transcription in Saccharomyces cerevisiae. Mol Cell Biol *7*, 679-686.

Close, D., Johnson, S.J., Sdano, M.A., McDonald, S.M., Robinson, H., Formosa,

T., and Hill, C.P. (2011). Crystal structures of the S. cerevisiae Spt6 core and C-terminal tandem SH2 domain. J Mol Biol *408*, 697-713.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev *11*, 2755-2766.

Connelly, S., and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev *2*, 440-452.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640-649.

Das, C., Tyler, J.K., and Churchill, M.E. (2010). The histone shuffle: histone chaperones in an energetic dance. Trends Biochem Sci *35*, 476-489.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol *319*, 1097-1113.

de Almeida, S.F., and Carmo-Fonseca, M. (2008). The CTD role in cotranscriptional RNA processing and surveillance. FEBS Lett *582*, 1971-1976.

De Koning, L., Corpet, A., Haber, J.E., and Almouzni, G. (2007). Histone chaperones: an escort network regulating histone traffic. Nat Struct Mol Biol *14*, 997-1007.

DeGennaro, C.M., Alver, B.H., Marguerat, S., Stepanova, E., Davis, C.P., Bahler, J., Park, P.J., and Winston, F. (2013). Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. Mol Cell Biol *33*, 4779-4792.

Denis, C.L., and Malvar, T. (1990). The CCR4 gene from Saccharomyces cerevisiae is required for both nonfermentative and spt-mediated gene expression. Genetics *124*, 283-291.

Diebold, M.L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J., and Romier, C. (2010a). The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIS, Elongin A and Med26. EMBO J *29*, 3979-3991.

Diebold, M.L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., and Romier, C. (2010b). Noncanonical tandem SH2 enables interaction of elongation factor Spt6 with RNA polymerase II. J Biol Chem *285*, 38389-38398.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of

the chromatin fiber. Science *306*, 1571-1573.

Duina, A.A. (2011). Histone chaperones Spt6 and FACT: Similarities and differences in modes of action at transcribed genes. Genet Res Int *2011*, 625210.

Duncan, K., Umen, J.G., and Guthrie, C. (2000). A putative ubiquitin ligase required for efficient mRNA export differentially affects hnRNP transport. Curr Biol *10*, 687-696.

Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. J Biol Chem *266*, 71-75.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev *113*, 8456-8490.

Elsasser, S.J. (2013). A common structural theme in histone chaperones mimics interhistone contacts. Trends Biochem Sci *38*, 333-336.

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P.W., Kobayashi, R., Blencowe, B.J., and Ingles, C.J. (2002). Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. RNA *8*, 1102-1111.

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A.*, et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol Cell Biol *24*, 3324-3336.

Estruch, F., Peiro-Chova, L., Gomez-Navarro, N., Durban, J., Hodge, C., Del Olmo, M., and Cole, C.N. (2009). A genetic screen in Saccharomyces cerevisiae identifies new genes that interact with mex67-5, a temperature-sensitive allele of the gene encoding the mRNA export receptor. Mol Genet Genomics *281*, 125-134.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A *73*, 1897-1901.

Formosa, T. (2012). The role of FACT in making and breaking nucleosomes. Biochim Biophys Acta *1819*, 247-255.

Froyen, G., Belet, S., Martinez, F., Santos-Reboucas, C.B., Declercq, M., Verbeeck, J., Donckers, L., Berland, S., Mayo, S., Rosello, M.*, et al.* (2012). Copy-number gains of HUWE1 due to replication- and recombination-based

rearrangements. Am J Hum Genet *91*, 252-264.

Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C., Bauters, M., Govaerts, K., Vandeleur, L., Van Esch, H.*, et al.* (2008). Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. Am J Hum Genet *82*, 432-443.

Gunjan, A., and Verreault, A. (2003). A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in S. cerevisiae. Cell *115*, 537-549.

Gurard-Levin, Z.A., Quivy, J.P., and Almouzni, G. (2014). Histone chaperones: assisting histone traffic and nucleosome dynamics. Annu Rev Biochem *83*, 487-517.

Hagenbuchle, O., Wellauer, P.K., Cribbs, D.L., and Schibler, U. (1984). Termination of transcription in the mouse alpha-amylase gene Amy-2a occurs at multiple sites downstream of the polyadenylation site. Cell *38*, 737-744.

Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M., and Martens, J.A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. Genes Dev *25*, 29-40.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev *12*, 357-369.

He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. Nature *495*, 481-486.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E.*, et al.* (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J *31*, 2784-2797.

Hirose, Y., Tacke, R., and Manley, J.L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. Genes Dev *13*, 1234-1239.

Hirschhorn, J.N., Brown, S.A., Clark, C.D., and Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev *6*, 2288-2298.

Ho, C.K., and Shuman, S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. Mol Cell *3*, 405-411.

Hondele, M., and Ladurner, A.G. (2011). The chaperone-histone partnership: for

the greater good of histone traffic and chromatin plasticity. Curr Opin Struct Biol *21*, 698-708.

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C., and Stutz, F. (2010). Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev *24*, 1927-1938.

Inoue, S., Hao, Z., Elia, A.J., Cescon, D., Zhou, L., Silvester, J., Snow, B., Harris, I.S., Sasaki, M., Li, W.Y.*, et al.* (2013). Mule/Huwe1/Arf-BP1 suppresses Ras-driven tumorigenesis by preventing c-Myc/Miz1-mediated down-regulation of p21 and p15. Genes Dev *27*, 1101-1114.

Isrie, M., Kalscheuer, V.M., Holvoet, M., Fieremans, N., Van Esch, H., and Devriendt, K. (2013). HUWE1 mutation explains phenotypic severity in a case of familial idiopathic intellectual disability. Eur J Med Genet *56*, 379-382.

Ivanovska, I., Jacques, P.E., Rando, O.J., Robert, F., and Winston, F. (2011). Control of chromatin structure by spt6: different consequences in coding and regulatory regions. Mol Cell Biol *31*, 531-541.

Johnson, S.J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S.L., and Hill, C.P. (2008). Crystal structure and RNA binding of the Tex protein from Pseudomonas aeruginosa. J Mol Biol *377*, 1460-1473.

Kan, P.Y., Lu, X., Hansen, J.C., and Hayes, J.J. (2007). The H3 tail domain participates in multiple interactions during folding and self-association of nucleosome arrays. Mol Cell Biol *27*, 2084-2091.

Kaplan, C.D., Holland, M.J., and Winston, F. (2005). Interaction between transcription elongation factors and mRNA 3'-end formation at the Saccharomyces cerevisiae GAL10-GAL7 locus. J Biol Chem *280*, 913-922.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.

Kaplan, C.D., Morris, J.R., Wu, C., and Winston, F. (2000). Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. Genes Dev *14*, 2623-2634.

Keegan, B.R., Feldman, J.L., Lee, D.H., Koos, D.S., Ho, R.K., Stainier, D.Y., and Yelon, D. (2002). The elongation factors Pandora/Spt6 and Foggy/Spt5 promote transcription in the zebrafish embryo. Development *129*, 1623-1632.

Kemble, D.J., McCullough, L.L., Whitby, F.G., Formosa, T., and Hill, C.P. (2015). FACT disrupts nucleosome structure by binding H2A-H2B with conserved peptide motifs. Mol Cell *60*, 294-306.

Kemble, D.J., Whitby, F.G., Robinson, H., McCullough, L.L., Formosa, T., and

Hill, C.P. (2013). Structure of the Spt16 middle domain reveals functional features of the histone chaperone FACT. J Biol Chem *288*, 10188-10194.

Kettenberger, H., Armache, K.J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol Cell *16*, 955-965.

Kim, D.H., and Koepp, D.M. (2012). Hect E3 ubiquitin ligase Tom1 controls Dia2 degradation during the cell cycle. Mol Biol Cell *23*, 4203-4211.

Kim, D.H., Zhang, W., and Koepp, D.M. (2012). The Hect domain E3 ligase Tom1 and the F-box protein Dia2 control Cdc6 degradation in G1 phase. J Biol Chem *287*, 44212-44220.

Kim, J., Guermah, M., and Roeder, R.G. (2010). The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. Cell *140*, 491-503.

Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. EMBO J *23*, 354-364.

Kok, F.O., Oster, E., Mentzer, L., Hsieh, J.C., Henry, C.A., and Sirotkin, H.I. (2007). The role of the SPT6 chromatin remodeling factor in zebrafish embryogenesis. Dev Biol *307*, 214-226.

Kornberg, R.D. (1977). Structure of chromatin. Annu Rev Biochem *46*, 931-954.

Kornberg, R.D. (2007). The molecular basis of eukaryotic transcription. Proc Natl Acad Sci U S A *104*, 12955-12961.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 Is an RNA polymerase II CTD phosphatase. Mol Cell *14*, 387-394.

Laskey, R.A., Honda, B.M., Mills, A.D., and Finch, J.T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature *275*, 416-420.

Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet *36*, 900-905.

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Curr Opin Genet Dev *21*, 175-186.

Lian, Z., Karpikov, A., Lian, J., Mahajan, M.C., Hartman, S., Gerstein, M., Snyder, M., and Weissman, S.M. (2008). A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation.

Genome Res *18*, 1224-1237.

Liu, J., Zhang, J., Gong, Q., Xiong, P., Huang, H., Wu, B., Lu, G., Wu, J., and Shi, Y. (2011). Solution structure of tandem SH2 domains from Spt6 protein and their binding to the phosphorylated RNA polymerase II C-terminal domain. J Biol Chem *286*, 29218-29226.

Liu, Y.X., Zhang, S.F., Ji, Y.H., Guo, S.J., Wang, G.F., and Zhang, G.W. (2012). Whole-exome sequencing identifies mutated PCK2 and HUWE1 associated with carcinoma cell proliferation in a hepatocellular carcinoma patient. Oncol Lett *4*, 847-851.

Liu, Z., Oughtred, R., and Wing, S.S. (2005). Characterization of E3Histone, a novel testis ubiquitin protein ligase which ubiquitylates histones. Mol Cell Biol *25*, 2819-2831.

Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proc Natl Acad Sci U S A *84*, 8306-8310.

Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1992). Initiation on chromatin templates in a yeast RNA polymerase II transcription system. Genes Dev *6*, 2282-2287.

Luger, K., and Hansen, J.C. (2005). Nucleosome and chromatin fiber dynamics. Curr Opin Struct Biol *15*, 188-196.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat Struct Mol Biol *17*, 1195-1201.

Machida, K., and Mayer, B.J. (2005). The SH2 domain: versatile signaling module and pharmaceutical target. Biochim Biophys Acta *1747*, 1-25.

Marino-Ramirez, L., Kann, M.G., Shoemaker, B.A., and Landsman, D. (2005). Histone structure and nucleosome stability. Expert Rev Proteomics *2*, 719-729.

Mattiroli, F., D'Arcy, S., and Luger, K. (2015). The right place at the right time: chaperoning core histone variants. EMBO Rep *16*, 1454-1466.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation

impairs termination factor recruitment to RNA polymerase II. Science *336*, 1723-1725.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol *17*, 1272-1278.

Maze, I., Noh, K.M., Soshnev, A.A., and Allis, C.D. (2014). Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nat Rev Genet *15*, 259-271.

McBryant, S.J., Klonoski, J., Sorensen, T.C., Norskog, S.S., Williams, S., Resch, M.G., Toombs, J.A., 3rd, Hobdey, S.E., and Hansen, J.C. (2009). Determinants of histone H4 N-terminal domain function during nucleosomal array oligomerization: roles of amino acid sequence, domain length, and charge density. J Biol Chem *284*, 16716-16722.

McCullough, L., Connell, Z., Petersen, C., and Formosa, T. (2015). The abundant histone chaperones Spt6 and FACT collaborate to assemble, inspect, and maintain chromatin structure in Saccharomyces cerevisiae. Genetics *201*, 1031-1045.

McDonald, S.M., Close, D., Xin, H., Formosa, T., and Hill, C.P. (2010). Structure and biological importance of the Spn1-Spt6 interaction, and its regulatory role in nucleosome binding. Mol Cell *40*, 725-735.

Meeks-Wagner, D., and Hartwell, L.H. (1986). Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. Cell *44*, 43-52.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature *430*, 223-226.

Morris, D.P., and Greenleaf, A.L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem *275*, 39935-39943.

Mosley, A.L., Hunter, G.O., Sardiu, M.E., Smolle, M., Workman, J.L., Florens, L., and Washburn, M.P. (2013). Quantitative proteomics demonstrates that the RNA polymerase II subunits Rpb4 and Rpb7 dissociate during transcriptional elongation. Mol Cell Proteomics *12*, 1530-1538.

Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J.*, et al.* (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. Science *342*, 1238724.

Myers, L.C., Gustafsson, C.M., Bushnell, D.A., Lui, M., Erdjument-Bromage, H.,

Tempst, P., and Kornberg, R.D. (1998). The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. Genes Dev *12*, 45-54.

Nava, C., Lamari, F., Heron, D., Mignot, C., Rastetter, A., Keren, B., Cohen, D., Faudet, A., Bouteiller, D., Gilleron, M.*, et al.* (2012). Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. Transl Psychiatry *2*, e179.

Neigeborn, L., Celenza, J.L., and Carlson, M. (1987). SSN20 is an essential gene with mutant alleles that suppress defects in SUC2 transcription in Saccharomyces cerevisiae. Mol Cell Biol *7*, 672-678.

Nishiwaki, K., Sano, T., and Miwa, J. (1993). emb-5, a gene required for the correct timing of gut precursor cell division during gastrulation in Caenorhabditis elegans, encodes a protein similar to the yeast nuclear protein SPT6. Mol Gen Genet *239*, 313-322.

Okazaki, I.M., Okawa, K., Kobayashi, M., Yoshikawa, K., Kawamoto, S., Nagaoka, H., Shinkura, R., Kitawaki, Y., Taniguchi, H., Natsume, T.*, et al.* (2011). Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. Proc Natl Acad Sci U S A *108*, 7920-7925.

Orphanides, G., LeRoy, G., Chang, C.H., Luse, D.S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. Cell *92*, 105-116.

Osheim, Y.N., Miller, O.L., Jr., and Beyer, A.L. (1985). RNP particles at splice junction sequences on Drosophila chorion transcripts. Cell *43*, 143-151.

Perales, R., Erickson, B., Zhang, L., Kim, H., Valiquett, E., and Bentley, D. (2013). Gene promoters dictate histone occupancy within genes. EMBO J *32*, 2645-2656.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A *90*, 7923-7927.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. Genes Dev *23*, 1247-1269.

Robinson, P.J., Fairall, L., Huynh, V.A., and Rhodes, D. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. Proc Natl Acad Sci U S A *103*, 6506-6511.

Rothbart, S.B., and Strahl, B.D. (2014). Interpreting the language of histone and DNA modifications. Biochim Biophys Acta *1839*, 627-643.

Saleh, A., Collart, M., Martens, J.A., Genereaux, J., Allard, S., Cote, J., and Brandl, C.J. (1998). TOM1p, a yeast hect-domain protein which mediates transcriptional regulation through the ADA/SAGA coactivator complexes. J Mol Biol *282*, 933-946.

Schuller, R., Forne, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-specific phosphorylation of RNA polymerase II CTD. Mol Cell *61*, 305-314.

Sekinger, E.A., Moqtaderi, Z., and Struhl, K. (2005). Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol Cell *18*, 735-748.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.

Shuman, S. (2001). Structure, mechanism, and evolution of the mRNA capping apparatus. Prog Nucleic Acid Res Mol Biol *66*, 1-40.

Singh, R.K., Kabbaj, M.H., Paik, J., and Gunjan, A. (2009). Histone levels are regulated by phosphorylation and ubiquitylation-dependent proteolysis. Nat Cell Biol *11*, 925-933.

Sogaard, T.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. J Biol Chem *282*, 14113-14120.

Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P., and Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science *344*, 376-380.

Spahr, H., Calero, G., Bushnell, D.A., and Kornberg, R.D. (2009). Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. Proc Natl Acad Sci U S A *106*, 9185-9190.

Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct analysis of phosphorylation sites on the Rpb1 C-Terminal domain of RNA polymerase II. Mol Cell *61*, 297-304.

Suh, H., Hazelbaker, D.Z., Soares, L.M., and Buratowski, S. (2013). The C-terminal domain of Rpb1 functions on other RNA polymerase II subunits. Mol Cell *51*, 850-858.

Sun, M., Lariviere, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A tandem

SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). J Biol Chem *285*, 41597-41603.

Swanson, M.S., and Winston, F. (1992). SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in Saccharomyces cerevisiae. Genetics *132*, 325-336.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol *14*, 1025-1040.

Thomas, J.O. (1999). Histone H1: location and role. Curr Opin Cell Biol *11*, 312-317.

Utsugi, T., Hirata, A., Sekiguchi, Y., Sasaki, T., Toh-e, A., and Kikuchi, Y. (1999). Yeast tom1 mutant exhibits pleiotropic defects in nuclear division, maintenance of nuclear structure and nucleocytoplasmic transport at high temperatures. Gene *234*, 285-295.

Utsugi, T., Toh-e, A., and Kikuchi, Y. (1995). A high dose of the STM1 gene suppresses the temperature sensitivity of the tom1 and htr1 mutants in Saccharomyces cerevisiae. Biochim Biophys Acta *1263*, 285-288.

Vanti, M., Gallastegui, E., Respaldiza, I., Rodriguez-Gil, A., Gomez-Herreros, F., Jimeno-Gonzalez, S., Jordan, A., and Chavez, S. (2009). Yeast genetic analysis reveals the involvement of chromatin reassembly factors in repressing HIV-1 basal transcription. PLoS Genet *5*, e1000339.

Wang, X., and Hayes, J.J. (2008). Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. Mol Cell Biol *28*, 227-236.

Williams, S.P., Athey, B.D., Muglia, L.J., Schappe, R.S., Gough, A.H., and Langmore, J.P. (1986). Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. Biophys J *49*, 233-248.

Winston, F., Chaleff, D.T., Valent, B., and Fink, G.R. (1984). Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics *107*, 179-197.

Woodcock, C.L., and Dimitrov, S. (2001). Higher-order structure of chromatin and chromosomes. Curr Opin Genet Dev *11*, 130-135.

Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes Dev *21*, 160-174.

Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev *22*, 3422-3434.

Zhao, X., D, D.A., Lim, W.K., Brahmachary, M., Carro, M.S., Ludwig, T., Cardo, C.C., Guillemot, F., Aldape, K., Califano, A.*, et al.* (2009). The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. Dev Cell *17*, 210-221.

CHAPTER 2

# CRYSTAL STRUCTURES OF THE *S. CEREVISIAE* SPT6 CORE AND C-TERMINAL TANDEM SH2 DOMAIN

Devin Close, Sean J. Johnson, Matthew A. Sdano,

Seth M. McDonald, Howard Robinson,

Tim Formosa, and

Christopher P. Hill

(2011)

Journal of Molecular Biology

408, 697-713



Note: See Chapter 1, Outline of Chapters, for a description of my contributions to this publication.

doi:10.1016/j.jmb.2011.03.002 J. Mol. Biol. (2011) 408, 697–713

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

# Crystal Structures of the *S. cerevisiae* Spt6 Core and C-Terminal Tandem SH2 Domain

**Devin Close[1], Sean J. Johnson[2], Matthew A. Sdano[1], Seth M. McDonald[1], Howard Robinson[3], Tim Formosa[1] and Christopher P. Hill[1]***

[1]*Department of Biochemistry, University of Utah, Salt Lake City, UT 84112-5650, USA*
[2]*Department of Chemistry and Biochemistry, Utah State University, Logan, UT 84322-0300, USA*
[3]*Department of Biology, Brookhaven National Laboratory, Upton, NY 11973, USA*

*Received 2 December 2010; received in revised form 24 February 2011; accepted 1 March 2011*
Available online 17 March 2011

***Edited by K. Morikawa***

***Keywords:***
protein structure; protein function; gene expression; crystallography

The conserved and essential eukaryotic protein Spt6 functions in transcription elongation, chromatin maintenance, and RNA processing. Spt6 has three characterized functions. It is a histone chaperone capable of reassembling nucleosomes, a central component of transcription elongation complexes, and is required for recruitment of RNA processing factors to elongating RNA polymerase II (RNAPII). Here, we report multiple crystal structures of the 168-kDa Spt6 protein from *Saccharomyces cerevisiae* that together represent essentially all of the ordered sequence. Our two structures of the ~900-residue core region reveal a series of putative nucleic acid and protein–protein interaction domains that fold into an elongated form that resembles the bacterial protein Tex. The similarity to a bacterial transcription factor suggests that the core domain performs nucleosome-independent activities, and as with Tex, we find that Spt6 binds DNA. Unlike Tex, however, the Spt6 S1 domain does not contribute to this activity. Crystal structures of the Spt6 C-terminal region reveal a tandem SH2 domain structure composed of two closely associated SH2 folds. One of these SH2 folds is cryptic, while the other shares striking structural similarity with metazoan SH2 domains and possesses structural features associated with the ability to bind phosphorylated substrates including phosphotyrosine. Binding studies with phosphopeptides that mimic the RNAPII C-terminal domain revealed affinities typical of other RNAPII C-terminal domain-binding proteins but did not indicate a specific interaction. Overall, these findings provide a structural foundation for understanding how Spt6 encodes several distinct functions within a single polypeptide chain.



**Corresponding author.* E-mail address: chris@biochem.utah.edu.
Present address: D. Close, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

Abbreviations used: CTD, C-terminal domain; DLD, death-like domain; dsDNA, double-stranded DNA; FA, fluorescence anisotropy; HhH, helix–hairpin–helix; HtH, helix–turn–helix; NSLS, National Synchrotron Light Source; OB, oligonucleotide–oligosaccharide binding; PDB, Protein Data Bank; pSer, phosphoserine; pTyr, phosphotyrosine; RNAPII, RNA polymerase II; SH2, Src homology 2; SSRL, Stanford Synchrotron Radiation Laboratory; WT, wild type; SeMet, selenomethionine substituted; NIH, National Institutes of Health.

0022-2836/$ - see front matter 

## Introduction

Gene expression in eukaryotes relies on a synergistic relationship between transcription, RNA processing, and chromatin structure.[1,2] The specific positioning, composition, and posttranslational modification of nucleosomes defines a code for chromatin-templated transcriptional regulation. Moreover, transcription is intimately tied to mRNA processing, surveillance, and export from the nucleus. This coordination relies on precise cooperation among many proteins, with Spt6 being remarkable for playing multifaceted roles in several distinct processes.

Spt6 (suppressor of Ty 6) was originally discovered in *Saccharomyces cerevisiae* as a gene that influences general transcription through manipulation of chromatin structure at upstream promoter elements.[3] Subsequently, Spt6 has been implicated in a variety of biological processes in organisms ranging from yeasts to human, including embryogenesis in zebrafish,[4] multiple stages of development in *Drosophila*,[5] gut morphogenesis in *Caenorhabditis elegans*,[6] signal transduction in mammals,[7,8] and pathogenesis of human immunodeficiency virus.[9,10] The broad utility of Spt6 stems from its ability to perform multiple functions as a histone chaperone, a transcription elongation factor, and a modulator of RNA transcript processing.

Spt6 is required for reassembly of nucleosomes in the wake of an elongating RNA polymerase II (RNAPII), a function that has profound regulatory effects at both intergenic and intragenic start sites.[11,12] Spt6 binds directly to histones and nucleosomes *in vitro*,[13,14] and these activities may contribute to the nucleosome reassembly function. In addition, Spt6 recruits the H3K36 methyltransferase Set2 to the transcription complex,[15] providing a link between the processes of transcription and histone modification. While its roles in modifying and reassembling nucleosomes indirectly influence the elongation rate, Spt6 also directly affects RNAPII, as it stimulates elongation on nucleosome-free DNA templates *in vitro*.[9,16] This role as an elongation factor independent of its effects on chromatin may also be significant *in vivo*, as knocking down Spt6 caused a decrease in the RNAPII elongation rate even in regions where the chromatin was considered to be permissive to transcription.[16] Yet another role as a modulator of transcript processing is indicated by the association of Spt6 with the Rrp6 subunit of the *Drosophila* exosome RNA processing complex[17] and by the requirement for Spt6 to prevent premature 3′ processing at cryptic polyadenylation signals upstream of the appropriate sites.[18] It has also been demonstrated that mammalian Spt6 can bind RNAPII C-terminal domain (CTD) phosphorylated at Ser2 by the P-TEFb kinase and that this interaction can subsequently promote recruitment of RNA processing/export factors such as REF1/Aly.[9,15] Binding to the phosphorylated RNAPII CTD is mediated by a Src homology 2 (SH2) domain that is located near the C-terminus of Spt6 and is conserved from yeast to human.[9] SH2 domains typically recognize phosphorylated tyrosine residues, are ubiquitous in metazoans, and are the primary recognition motif in phosphorylation-mediated signal transduction cascades.[19] Strikingly, the Spt6 SH2 domain is the only SH2 domain predicted to occur in the yeast

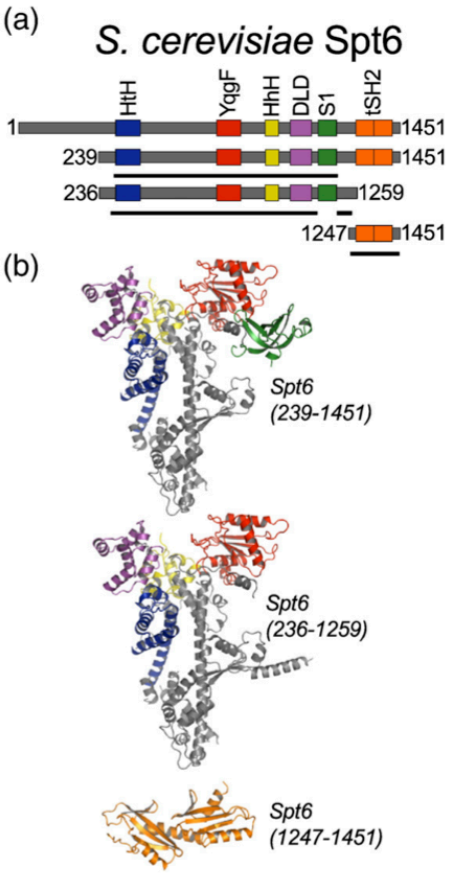



**Fig. 1.** Spt6 structures. (a) Schematic representation of full-length (top) Spt6 and constructs used for crystallization. The three crystallized constructs (239–1451, 236–1259, and 1247–1451) are shown below, with black continuous lines indicating regions of the protein constructs visible in each of the crystal structures. (b) Three independently determined Spt6 crystal structures colored by domain as in (a).

proteome.[20] Spt6 therefore participates in a wide range of functions affecting transcription, with each activity requiring different subsets of its multiple distinct functional domains.

We have determined multiple crystal structures of Spt6 from *S. cerevisiae* and find that, consistent with the range of functional domains inferred from previous studies, it comprises a series of structural domains whose homologs are known to function in nucleic acid binding and/or protein–protein interactions. The core of the structure comprises several recognizable structural motifs and, in composite, resembles the bacterial transcription factor Tex.[21] A C-terminal region that is tethered to the core by a flexible linker adopts a novel tandem SH2 domain comprising two closely associated SH2 folds, one of which corresponds to the previously predicted SH2 domain of Spt6 and contains many of the standard binding determinants characteristic of this family, while the other lacks these features but contributes to a putative specificity pocket of the more canonical SH2 domain. Our structure of the Spt6 tandem SH2 domain resembles two recently reported homologous Spt6 structures.[22,23] We also show that the Spt6 core domain has DNA-binding activity, and we examine the interaction between the Spt6 tandem SH2 domain and RNAPII-derived peptides for evidence of a phosphorylation-dependent interaction with the CTD.

## Results and Discussion

### Crystal structures of Spt6(236–1259), Spt6(239–1451), and Spt6(1247–1451)

We have determined three crystal structures that together comprise the entire ordered region of the 1451-residue *S. cerevisiae* Spt6 protein (Figs. 1–3). Based on these structures and sequence analysis, Spt6 residues 1–297, 456–464, 485–500, 562–566, 1003–1008, 1211–1217, and 1441–1451 are likely to

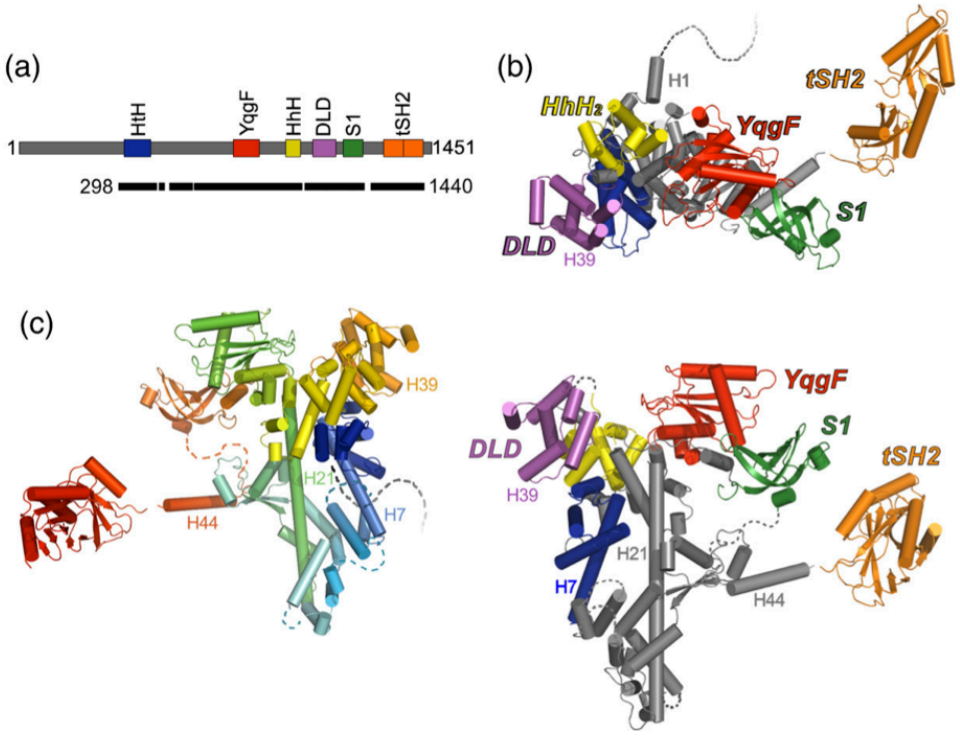



**Fig. 2.** Composite model of Spt6. (a) Schematic model of the Spt6 protein as in Fig. 1a. Black bar indicates segments of the protein represented by the composite model. (b) Two views of the composite model of Spt6 colored by domain. Approximately two hundred C-terminal residues (tSH2 domain) are expected to be highly mobile with respect to the core. Secondary structure elements mentioned in the text are labeled. Broken lines represent regions of the Spt6 protein that are not visible in our structures and are likely to be disordered, including the first 297 residues. (c) Composite model of Spt6 colored from the N- to C-terminus (blue to red). View orientation is rotated 180° from that of the lower image in (b).

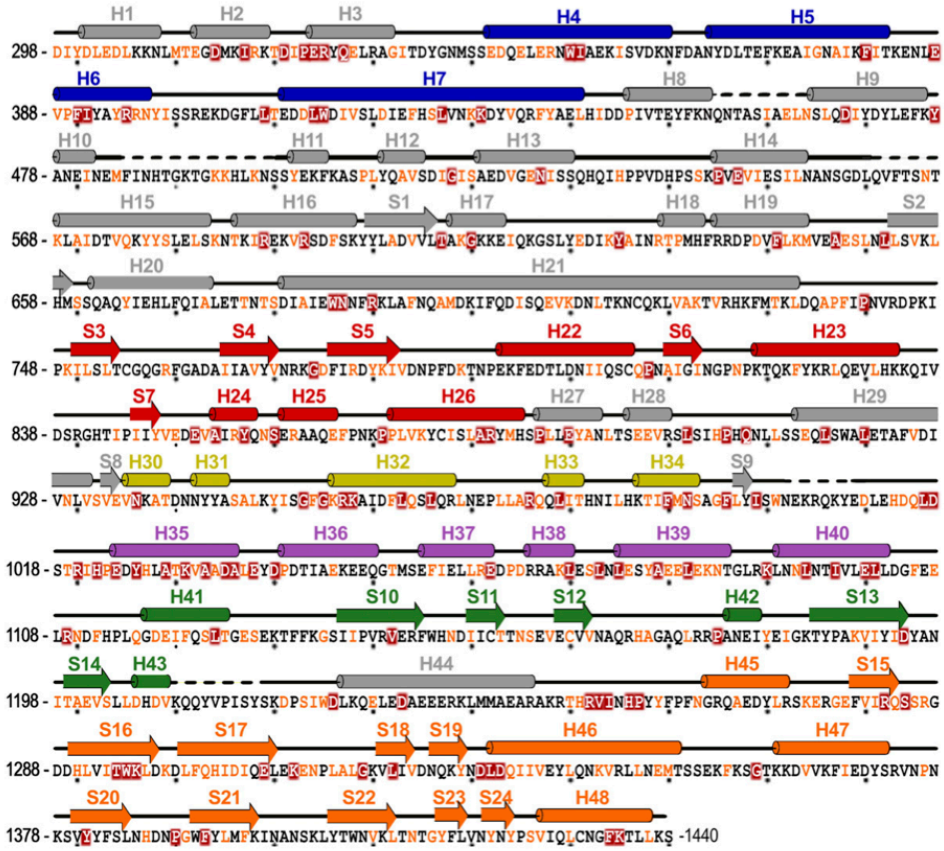



**Fig. 3.** Spt6 sequence. Spt6 sequence present in the composite model with corresponding secondary structure elements colored by domain as in Figs. 1 and 2. Coloring of sequence represents degree of conservation [dark-red background, invariant; orange font, conserved] in an alignment (see Supplemental Fig. S1) of proteins from *S. cerevisiae*, *Schizosaccharomyces pombe*, *C. elegans*, *Drosophila melanogaster*, *Danio rerio*, and *Homo sapiens*. Alignment was performed using T-coffee.[24] Broken lines indicate regions of disorder that are not included in the model(s).

be disordered in the full-length protein, at least in the absence of binding partners. Spt6 displays multiple recognizable structural domains whose homologs have been implicated in binding of nucleic acids or proteins, although conservation at the sequence level is low, and only three of these domains [$(HhH)_2$, YqgF, and S1] in the Spt6 core were predicted from the sequence.[25]

Full-length Spt6 expressed poorly in *Escherichia coli* and did not yield crystals. In contrast, two different Spt6 constructs that lacked the first ~235 residues and either lacked the C-terminal 192 residues [Spt6 (236–1259)] or extended to the C-terminus [Spt6 (239–1451)] expressed well, and the resulting proteins crystallized in different space groups (Table 1). The Spt6(236–1259) structure was determined by two-wavelength anomalous diffraction using data collected to 2.7 Å from crystals of selenomethionine-substituted (SeMet) protein and was refined against 2.6-Å native data to $R/R_{free}$ values of 22.4%/26.5%. The Spt6(239–1451) structure was determined by molecular replacement using the Spt6(236–1259) structure as a search model and refined against 3.3-Å data to $R/R_{free}$ values of 26.5%/30.8%.

Together, Spt6(236–1259) and Spt6(239–1451) display an ordered structure for the Spt6 core (residues 298–1248), which is primarily helical (54.6% helical, 8.3% strand, 37.1% coil), has overall dimensions of 110 Å × 77 Å × 36 Å, and is quite similar between the two structures (RMSD = 1.2 Å over 747 out of 763 pairs of $C^{\alpha}$ atoms that are ordered in both structures). The core is built around an ~80-Å-long central helix (H21, 680–733), with the rest of the

**Table 1.** Crystallization conditions

| | Spt6(236–1259) | Spt6(236–1259) | Spt6(239–1451) | Spt6(1247–1451) | Spt6(1247–1451) |
|---|---|---|---|---|---|
| Crystal | SeMet | Native | Native | SeMet | Native |
| Protein solution[a] | 15 mM Tris, 5% glycerol | 15 mM Tris, 5% glycerol | 50 mM Tris, 10% glycerol, 5 mM β-mercaptoethanol | 15 mM Tris, 5% glycerol | 15 mM Tris, 5% glycerol |
| Protein concentration (mg/ml) | 9 | 8 | 10 | 26 | 26 |
| Reservoir solution | 50 mM 4-morpholineethanesulfonic acid (pH 6.0), 13% methyl-2,4-pentanediol, 0.1 M KCl, 5 mM $MgCl_2$ | 50 mM $Mg(HCO_2)_2$, 10% glycerol | 0.1 M Na–4-morpholineethanesulfonic acid (pH 6.5), 13% PEG 4000, 0.2 M $MgCl_2$, 12% glycerol | 22% PEG 4000, 0.3 M $(NH_4)_2SO_4$ | 0.1 M Tris (pH 5.5), 25% PEG 3350, 0.2 M $(NH_4)_2SO_4$ |
| Drop ratio (μl:μl), protein:reservoir | 2:2 | 3:1 | 3:1.5 | 2: 2 | 1:2 |
| Additive | — | 0.6 μl 1.0 M NDSB-256 | — | — | — |
| Temperature (°C) | 13 | 13 | 13 | 21 | 21 |
| Cryo solution[b] | 30% methyl-2,4-pentanediol | 35% glycerol | 30% glycerol | 30% glycerol | 20% glycerol |

PEG, polyethylene glycol.
[a] All protein solutions were in 100 mM NaCl and Tris, pH 7.5 (defined at room temperature).
[b] Cryoprotectant solutions were made the same as for the reservoir solution, but including glycerol.

protein chain wrapping around this helix at both ends to give an overall V shape. Additionally, Spt6 (239–1451) displayed interpretable density for the S1 domain (residues 1128–1210), whereas the corresponding density in Spt6(236–1259) was poorly defined, and the S1 domain was not included in this refined structure. As discussed below, we favor a model in which the S1 domain is highly mobile in solution. Interestingly, the 240 C-terminal residues of Spt6(239–1451), encompassing all of the residues absent from the shorter Spt6(236–1259) construct, are not visible in the electron density. The Spt6(239–1451) crystals have an estimated solvent content of 60% with cavities that could accommodate the 240 C-terminal residues, which suggests that this region is tethered by a flexible linker that is mobile both in solution and in the Spt6(239–1451) crystals.

Guided by secondary structure predictions and limited proteolysis experiments (data not shown), we expressed and crystallized the Spt6 C-terminal region, Spt6(1247–1451). This structure was determined by the selenomethionine single-wavelength anomalous diffraction method using 2.7-Å-resolution data and refined to $R/R_{free}$ values of 20.7%/25.4%. A second native crystal form yielded 2.1-Å-resolution data, and this model was refined to $R/R_{free}$ values of 17.9%/21.2%. The SeMet and native proteins crystallized in different space groups with one and four molecules in the asymmetric unit, respectively (Table 2). Surprisingly, we found that the Spt6 C-terminal region comprises not one SH2 domain as anticipated,[20,28] but two SH2 folds that are packed tightly against each other to form a tandem SH2 domain composed of N-terminal (nSH2, residues 1250–1353) and C-terminal (cSH2, residues 1354–1440) folds. While this manuscript was in the final stages of preparation, equivalent tSH2 structures that overlap closely with our refined model were reported for Spt6 homologs from *Candida glabrata*[22] (RMSD = 1.0 Å over 184 $C^{\alpha}$ atoms with 87% sequence identity) and *Antonospora locustae*[23] (RMSD = 1.6 Å over 164 $C^{\alpha}$ atoms with 24% sequence identity).

## N-terminal region

The first ~300 amino acids of Spt6 are extremely acidic, with an overall charge of −62 and a predicted p*I* of 4.3, and are also predicted to be disordered.[29] This is consistent with our Spt6(236–1259) and Spt6 (239–1451) structures, which lacked discernible density prior to residue 298. Despite the lack of inherent order, this region of Spt6 is functionally important. Residues 239–268 bind the essential Spn1/Iws1 protein and overlap with residues required for nucleosome binding.[14] Furthermore, the *spt6-1002* allele, a deletion of residues 2–122, displays synthetic lethality with deletion of the gene encoding the transcription factor Paf1.[18]

### Helix–turn–helix domain

Residues 336–442 resemble a DNA-binding helix–turn–helix (HtH) motif, as seen in transcription factors such as c-Myb and the bacterial sigma factors.[30] Nevertheless, the structure is not consistent with binding DNA in a canonical manner; binding of Spt6 H7 in the major groove of DNA in the manner predicted for a canonical HtH domain would cause steric clashes between bound DNA and the rest of the structure. It is possible that Spt6 undergoes conformational changes upon binding DNA or that the Spt6 HtH domain serves as a protein–protein interaction motif, as occurs with members of the PWI subgroup of HtH domains.[31] The Spt6 HtH overlaps with the U4/U6 ribonucleoprotein Prp3 PWI domain [RMSD of 2.7 Å, 69 $C^{\alpha}$, Protein Data Bank (PDB) code 1x4q] and the Nab2 PWI domain (RMSD of 2.9 Å, 73 $C^{\alpha}$, PDB code 2v75), conserves the eponymous PWI motif as NWI (Asn349-Trp350-Ile351), and could utilize the equivalent protein-binding surface without invoking a conformational change in Spt6.

### YqgF homologous domain

The Spt6 YqgF domain (residues 735–887) resembles members of the YqgFc superfamily, such as the *E. coli* protein YqgF and the RuvC class of Holliday junction resolvases.[25] The alignment is especially close with *E. coli* RuvC (RMSD of 2.9 Å, 117 $C^{\alpha}$, PDB code 1hjr). Despite this similarity, the putative Spt6 YqgF catalytic site lacks carboxylate side chains that are critical for coordinating magnesium ions

**Table 2.** Data collection and refinement statistics

| | Se-Spt6 (236–1259) peak | Se-Spt6 (236–1259) inflection | Spt6 (236–1259) | Spt6 (239–1451) | Se-Spt6 (1247–1451) peak | Spt6 (1247–1451) |
|---|---|---|---|---|---|---|
| *Data collection* | | | | | | |
| Space group | $P2_12_12_1$ | | $P2_12_12_1$ | $P3_121$ | $I4_122$ | $P2_12_12_1$ |
| Unit cell dimensions (Å) | $a$=114.0 | | $a$=115.1 | $a$=118.7 | $a$=97.5 | $a$=78.8 |
| | $b$=116.4 | | $b$=116.2 | $b$=118.7 | $b$=97.5 | $b$=105.2 |
| | $c$=122.7 | | $c$=117.4 | $c$=214.4 | $c$=132.4 | $c$=119.5 |
| Molecules per asymmetric unit | 1 | | 1 | 1 | 1 | 4 |
| Solvent content (%) | 63.7 | | 62.4 | 60.0 | 63.5 | 53.1 |
| Beamline[a] | SSRL 11-1 | SSRL 11-1 | SSRL 7-1 | NSLS X29 | SSRL 9-1 | Home |
| Wavelength (Å) | 0.97886 | 0.97922 | 0.97773 | 1.10000 | 0.97908 | 1.54178 |
| Resolution (Å) | 50–2.7 | 50–2.7 | 32–2.6 | 46–3.3 | 35–2.7 | 30–2.1 |
| High-resolution shell (Å) | (2.8–2.7) | (2.8–2.7) | (2.7–2.6) | (3.42–3.3) | (2.8–2.7) | (2.18–2.1) |
| No. of unique reflections | 83,751 | 82,946 | 49,595 | 27,025 | 8481 | 58,489 |
| No. of total reflections | 261,886 | 257,334 | 434,557 | 182,336 | 649,419 | 258,498 |
| Mean $I/\sigma_I$ | 22.1 (2.4) | 22.1 (2.1) | 29.8 (4.3) | 39.8 (3.3) | 25.1 (3.7) | 16.8 (3.2) |
| Completeness (%) | 95.8 (75.8) | 94.7 (70.1) | 99.8 (99.9) | 99.5 (99.9) | 92.2 (74.5) | 99.5 (99.2) |
| $R_{sym}$ (%)[b] | 4.8 (35.7) | 4.7 (38.3) | 6.5 (43.9) | 5.8 (56.9) | 7.0 (29.6) | 6.6 (47.8) |
| *Refinement* | | | | | | |
| $R_{cryst}$[c]/$R_{free}$[d] (%) | | | 22.4/26.5 | 26.5/30.8 | 20.7/25.4 | 17.9/21.2 |
| No. of non-H atoms | | | | | | |
| Protein | | | 6788 | 6876 | 1575 | 6988 |
| Solvent | | | 48 | 0 | 19 | 491 |
| $\langle B \rangle$ (Å$^2$) | | | 103.7 | 167.6 | 74.7 | 44.2 |
| RMSD bond lengths (Å) | | | 0.006 | 0.013 | 0.008 | 0.004 |
| RMSD bond angles (°) | | | 0.895 | 1.59 | 1.12 | 0.757 |
| Ramachandran outliers (%) | | | 0.0 | 0.8 | 0.0 | 0.40 |
| Ramachandran favored (%) | | | 95.6 | 93.6 | 97.9 | 97.5 |
| Rotamer outliers (%) | | | 0.94 | 0.3 | 0.6 | 1.4 |

Values in parentheses correspond to the high-resolution shell.
Refinement statistics were determined by PHENIX[26] and MolProbity.[27]
[a] Data were collected at the SSRL, the NSLS, or on a Rigaku MicroMax-007HF rotating anode X-ray generator with a copper anode and VariMax confocal optics and a Rigaku R-AXIS IV image plate detector (home).
[b] $R_{sym} = (\sum |(I - \langle I \rangle)|)/(\sum I)$, where $\langle I \rangle$ is the average intensity of multiple measurements.
[c] $R_{cryst} = (\sum |F_{obs} - F_{calc}|)/(\sum |F_{obs}|)$.
[d] $R_{free}$ is the $R_{cryst}$ based on ~1000 (at least 10%) of the reflections that were excluded from refinement.

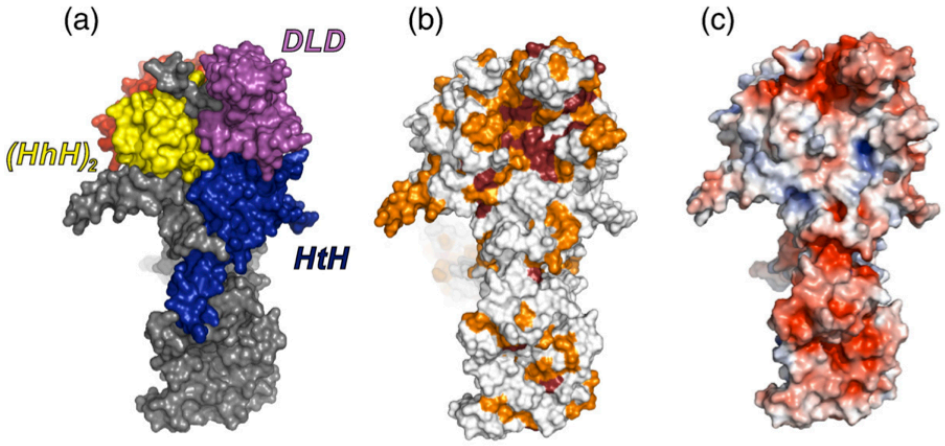



**Fig. 4.** The most conserved surface of the Spt6 core. (a) View of the Spt6 core showing the interface between HtH, DLD, and $(HhH)_2$ domains. (b) Same orientation as (a) but colored by conservation to illustrate the high level of surface conservation at the intersection of these domains, especially on the DLD. Coloring represents degree of conservation as described in Fig. 3 and Supplemental Fig. S1. (c) Same as (b) but colored by electrostatic potential (−5 to +5 $kT/e$).

that mediate phosphodiester bond hydrolytic cleavage.[32] Thus, it does not appear that the Spt6 YqgF fold is capable of nuclease activity using a catalytic mechanism similar to that of RuvC or related RNase H-fold nucleases.

### Helix–hairpin–helix domain

Residues 933–1002 form two consecutive helix–hairpin–helix (HhH) motifs that pack together through highly conserved hydrophobic residues at an ~90° angle to form a $(HhH)_2$ domain that resembles known double-stranded DNA (dsDNA) binding domains.[33] These include proteins such as *E. coli* RNA polymerase α CTD (RMSD of 2.3 Å, 56 $C^{\alpha}$, PDB code 1lb2) and the Holliday junction-binding protein RuvA (RMSD of 2.2 Å, 60 $C^{\alpha}$, PDB code 1bvs). The first Spt6 HhH represents a characteristic HhH motif in the relative angle of the antiparallel helices and the presence of the Gly-hydrophobic-Gly motif within the hairpin loop.[33] The second HhH motif is more variant, as also observed in other $(HhH)_2$ domains, including DNA polymerase β and 5′ to 3′ exonucleases. Though

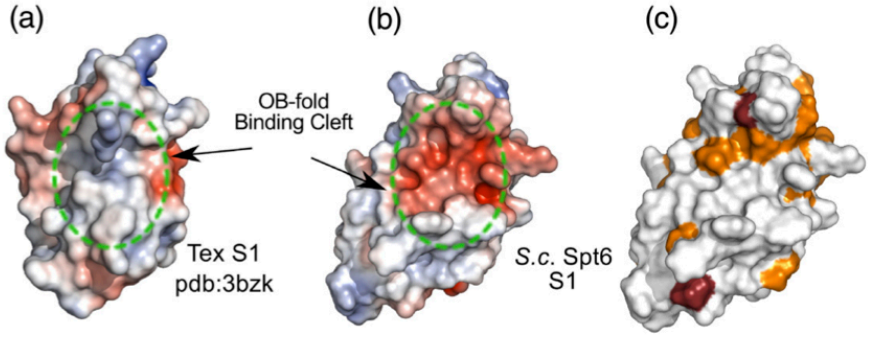



**Fig. 5.** Surface representations of the Spt6 and Tex S1 domains. (a) Electrostatic surface representation (−5 to +5 $kT/e$) of the S1 domain from *Pseudomonas aeruginosa* Tex (PDB code 3bzk) with the position of the nucleic acid binding OB-fold cleft approximated by the circle with the dotted green line. (b) Electrostatic surface representation (−5 to +5 $kT/e$) of the *S. cerevisiae* Spt6 S1 domain in the same orientation as in (a) showing a clustering of negative charge in the putative OB-fold binding cleft. (c) *S. cerevisiae* Spt6 S1 domain in the same orientation as (b) but colored by conservation in the same color scheme as in Fig. 3 to illustrate the low level of conservation within the region equivalent to the binding cleft of canonical S1 domains.

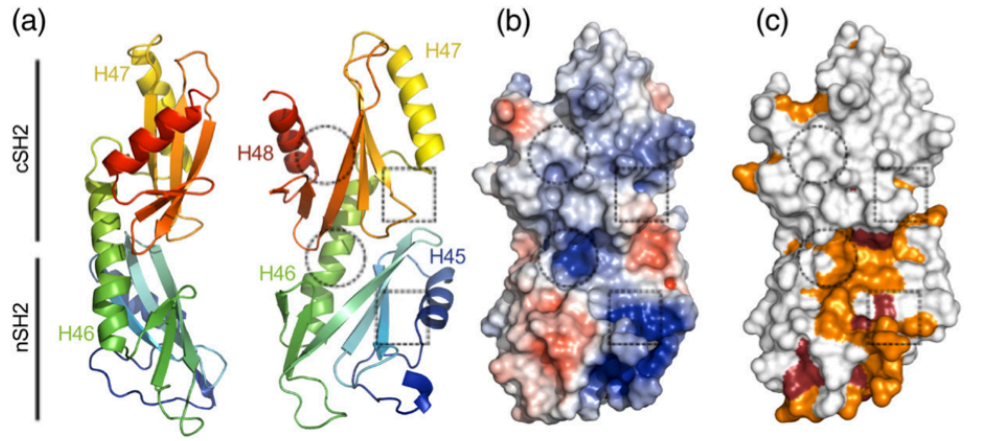



**Fig. 6.** The *S. cerevisiae* Spt6 tSH2 domain. (a) Two views of a cartoon representation of the tSH2 domain colored from the N- to C-terminus (blue to red; residues 1247–1440 are shown). Dotted squares and circles show the approximate positions of the canonical SH2 domain pTyr and specificity pockets, respectively. (b) Surface representation colored by electrostatic potential surface (−5 to +5 $kT/e$). (c) Same as (b) but colored by residue conservation in the same color scheme as in Figs. 3 and 4.

$(HhH)_2$ folds are primarily found in proteins that interact with DNA, they also occur in proteins that mediate protein–protein interactions, such as the sterile α motif proteins.[34] Notably, the first 28 ordered residues (~298–325) of Spt6 wrap around the $(HhH)_2$ domain in a fashion that would occlude binding of a canonical $(HhH)_2$ domain to a dsDNA ligand, although this interaction could be transient. The absence of corresponding density for H1 and part of H2 in the initial SeMet Spt6(236–1259) and Spt6(239–1451) maps suggests that these N-terminal residues might adopt a different conformation to allow binding of a physiological partner *in vivo*.

### Death-like domain

Residues 1019–1104 form a prominent lobe of the structure that resembles members of the death domain superfamily. Death domains typically serve as recognition modules in proteins that assemble and activate inflammatory and apoptotic complexes.[35] The Spt6 death-like domain (DLD) maintains the characteristic overall topology of death domains, consisting of a six-helix bundle with three stacked antiparallel helices but with an additional helix inserted between the final two helices of the bundle (H39 in Figs. 2 and 3). Spt6 aligns reasonably well with several known death domain superfamily proteins, including the caspase-2-activating PIDDosome PIDD protein subunit component (RMSD of 3.0 Å, 60 $C^{\alpha}$, PDB code 2of5). Although it is unlikely that the Spt6 DLD functions in an apoptotic process in yeast, its prominent location and the observation that it displays the most highly conserved region of the Spt6 surface suggest that it mediates important intermolecular interactions (Fig. 4).

### S1 domain

A mostly unstructured linker of 15 residues leads to the S1 domain (residues 1129–1219), which adopts the canonical S1/oligonucleotide–oligosaccharide binding (OB)-fold of a β-barrel composed of two three-stranded β-sheets where strand 1 (S10) is shared by both sheets.[36] Despite the structural similarity, Spt6 lacks the typical S1 binding cleft residues that are important for binding nucleic acids.[36] In addition, the predicted electrostatic potential surface does not appear conducive to nucleic acid binding, shows a low level of conservation, and, as discussed below, is not required for dsDNA binding (Fig. 5). This is in contrast to the distantly related bacterial Tex protein, which loses its capacity to bind DNA or RNA in the absence of the S1 domain.[21] OB folds are used to bind partners other than nucleic acids, including oligosaccharides and proteins;[36,37] thus, it remains possible that the Spt6 S1 domain is used for an important interaction that does not involve nucleic acids.

### Tandem SH2 domain

The S1 domain is followed by an unstructured ~10-residue segment and an ~30-Å helix (H44; 1227–1247) that buries ~440 $Å^2$ of accessible surface area against the core in the Spt6(236–1259) structure (Fig. 2). While some density is present for H44 in the

Spt6(239–1451) maps, this region is too disordered for reliable model building, which may indicate that this interface is not always formed in solution. H44 links the core to a tandem SH2 domain (tSH2; residues 1250–1440) that comprises N-terminal (nSH2; residues 1250–1353) and C-terminal (cSH2; 1353–1440) folds that associate through an ~800-Å$^2$ interface to form a single structural unit (Fig. 6). Both nSH2 and cSH2 conform to the standard SH2 domain fold (standard SH2 nomenclature in parenthesis) of an N-terminal helix (αA), a central three-stranded β-sheet (βB–βD), a small two-strand extension to the β-sheet (βE–βF), and a second helix (αB).[38] Intervening loops are labeled based on their relative position between these elements (e.g., the BC loop connects the βB and βC strands). The Spt6 nSH2 and cSH2 superimpose well with each other (RMSD of 2.1 Å, 69 C$^\alpha$; Fig. 7) and with the multitude of other characterized SH2 domains, such as those from v-Src kinase (nSH2: RMSD of 2.0 Å, 82 C$^\alpha$; cSH2: RMSD of 2.1 Å, 80 C$^\alpha$, PDB code 1sps; Fig. 7) and Nck2 (nSH2: RMSD of 2.1 Å, 83 C$^\alpha$; cSH2: RMSD of 1.9 Å, 73 C$^\alpha$, PDB code 2cia).

The relative orientation of the two SH2 folds that comprise the Spt6 tSH2 domain is unlike that of previously reported tandem SH2 domains from

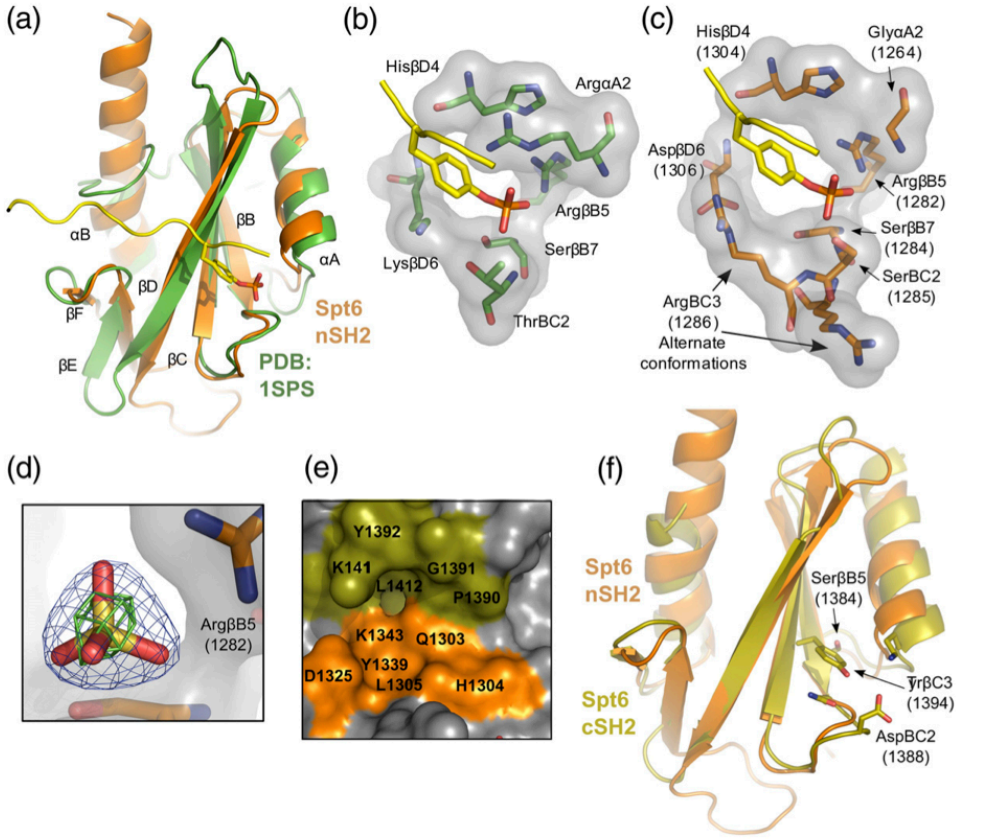



**Fig. 7.** tSH2 binding pockets. (a) Overlay of Spt6 nSH2 (orange) and the v-Src Kinase SH2 domain (green; PDB code 1sps) bound to a pTyr ligand (yellow). Secondary structure elements are labeled based on the standard SH2 nomenclature.[38] (b) Detailed view of the pTyr binding pocket of 1sps. Residues contributing to the coordination of the pTyr ligand are shown. (c) Same as (b) but for the Spt6 nSH2, with the pTyr peptide from 1sps positioned after the overlap on the SH2 protein domains. (d) Electron density for a sulfate bound in the tSH2 crystal structures. Blue density represents the $2mF_o - DF_c$ map contoured at 2.0 σ, and green density represents an anomalous difference Fourier map contoured at 3.0 σ. (e) The putative nSH2 specificity pocket. Residues from both nSH2 (orange) and cSH2 (olive) line the nSH2 specificity pocket. (f) Alignment of the Spt6 nSH2 (orange) and cSH2 (olive) folds. Residues that protrude into the typical location of the pTyr binding pocket of the cSH2 fold are shown.

other proteins. The αB helix of nSH2 undergoes an ~20° kink where the αB helix would end in a canonical SH2 domain and extends along the backside of the cSH2 fold to form an extensive hydrophobic packing interface with the cSH2 central β-sheet. The relative orientation of nSH2 and cSH2 folds therefore appears to be constrained, consistent with the observation that superposition of the five crystallographically independent tSH2 domains observed in our various crystal forms [one in SeMet Spt6(1247–1451) and four in native Spt6(1247–1451)] indicates a maximum relative rotation of ~8° between nSH2 and cSH2 folds.

### Implications for ligand binding by the tSH2 domain

The extent to which Spt6 nSH2 corresponds to a prototypic SH2 domain is seen in a comparison with v-Src (PDB code 1sps) (Fig. 7a–c). The primary determinants of phosphate binding are preserved in nSH2, whereas the positively charged Arg/Lys side chains of classical SH2 domains that flank the aromatic ring of phosphotyrosine (pTyr) ligands are absent from their usual positions. Conserved residues include the consensus FLVRES (FVIRQS, 1279–1284 in Spt6) sequence motif that contributes the phosphate-coordinating Arg1282 and Ser1284 side chains (Spt6 numbering). Moreover, the following Ser1285 Spt6 side chain is also well positioned to hydrogen bond the phosphate and functionally substitute the ThrBC2 of the Src SH2 domain. Other positions within SH2 domains that are important for binding phosphate include HisβD4 (Spt6 H1304) and Ser/ThrβC4 (Spt6 Thr1294), whose side chains hydrogen bond the Arg1282 side chain in an optimum orientation for phosphate binding and, in cognate SH2–ligand complexes, also form a main chain to main chain hydrogen bond with the ligand residue following the pTyr.

Classical SH2 domains typically have a basic residue at the αA2 position that binds against one side of the tyrosine ligand aromatic ring where it forms an amino–aromatic interaction with the π ring and also hydrogen bonds with both the phosphate and the tyrosine main-chain carbonyl. In contrast, the Spt6 nSH2 has a glycine at this position, G1264, which can make none of the same ligand interactions. This does not argue strongly against binding of pTyr by Spt6, however, because a number of other SH2 domains that bind pTyr ligands have a variety of substitutions at this position, including the PTPN11/SHPTP2/Syp phosphatase (PDB code 1ayc) that, like Spt6, has a Gly at αA2 and is known to bind a pTyr-containing peptide.[39] On the other side of a canonical pTyr ligand side chain, classical SH2 domains typically have a basic residue in the βD6 position. This residue is a lysine in Src but is an aspartate (D1306) in Spt6. Interestingly, in three of the five crystallographically independent Spt6 tSH2 molecules in our structures, the space typically occupied by the basic βD6 side chain is filled by R1286 in the BC3 position (Fig. 7c). Thus, Spt6 retains the ability to provide a positively charged basic group in this position, consistent with the potential to bind pTyr.

Ammonium sulfate was present in the crystallization solutions for both native and SeMet tSH2 domain structures, and all five crystallographically independent molecules displayed a sulfate ion at the putative nSH2 phosphate binding site, where it forms hydrogen bonds with R1282, S1284, and S1285 in the same manner as the phosphate of pTyr–SH2 complexes (Fig. 7d). Assignment of the density as sulfate was confirmed in anomalous difference Fourier maps for the native data, which showed peaks that were similar in size to those of cysteine and methionine sulfur atoms for some of the sulfates. This further suggests that nSH2 binds a phosphorylated ligand and that it might accommodate a pTyr side chain in a suboptimal binding pocket.

Typical SH2 domain ligands bind through a two-prong mechanism that, in addition to binding pTyr, also involves binding of the three side-chain residues C-terminal to the pTyr into the "specificity pocket."[38] Binding partner preference is typically defined in the specificity pocket by the BG and EF loops and the βD3 and βD5 residues, which usually favor binding of hydrophobic residues. In contrast, the Spt6 nSH2 fold predominantly displays charged and polar residues in this site, and there is no BG loop due to the extension of nSH2 αB to the cSH2 fold. Instead, cSH2 residues such as the DE loop and a βD side chain (K1411) protrude into the pocket, forming the top portion of the nSH2 specificity pocket (Fig. 7e). This indicates that, if the nSH2 fold binds substrate in the typical "two-pronged" manner common to SH2 domains, the cSH2 fold would make significant contributions to binding.

In contrast to nSH2, the cSH2 fold appears to be cryptic and unlikely to bind a phosphorylated ligand because residues critical for phosphate binding are substituted to display a very different chemical environment, and the Y1394 side chain fills the space where a phosphate would typically bind (Fig. 7f). Moreover, the region of the specificity pocket lacks even a shallow depression, as it is filled by bulky, aromatic side chains (F1397, Y1406, W1408, and F1434). Thus, in contrast to nSH2 and consistent with the lack of sequence conservation at cSH2 (Fig. 6c), it seems unlikely that cSH2 binds ligands in a manner reminiscent of SH2 domains.

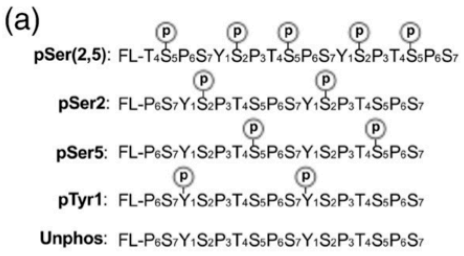
### Binding of Spt6 tSH2 domain with phosphorylated peptides

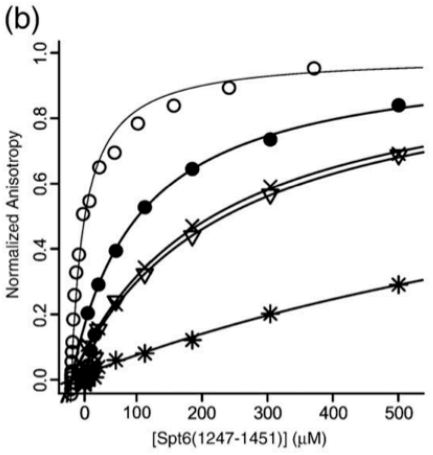
The human Spt6 CTD has been reported to bind the heptad repeat sequences of the mammalian

RNAPII large subunit when the RNAPII CTD is treated with P-TEFb, a kinase that phosphorylates Ser2 during transcriptional elongation.[9] In order to investigate this interaction more quantitatively, we used fluorescence anisotropy (FA) to measure binding of *S. cerevisiae* Spt6(1247–1451) to di-heptad repeat peptides representing various phosphoisoforms of the RNAPII CTD (Fig. 8). Peptides tested include sequences representing phosphoserine (pSer) 2, pSer5, pTyr1, and pSer(2,5). pSer5 and pSer(2,5) peptides were included to test specificity and because these modifications also occur on the RNAPII heptad repeats. A pTyr1 peptide was included because SH2 domains typically bind pTyr peptides and this modification occurs in mammals,[40] although it has not been reported to occur in yeast.

All of the peptides assayed bound with affinities in the range of ~20–250 μM, which is similar to the affinity of other RNAPII CTD interactions with isolated binding domains,[41] but is ~10- to 100-fold weaker than is typically found for the interaction of SH2 domains with pTyr ligands.[42] Of the ligands we assayed, the pSer(2,5) peptide bound Spt6(1247–1451) with the highest affinity (23 μM). This may indicate that RNAPII CTD sequences phosphorylated on both Ser2 and Ser5 are the authentic *in vivo* ligands for the yeast Spt6 tSH2 domain, which would be consistent with the reports that Ser2-phosphorylated sequences are preferred[9] and that localization of Ser2 and Ser2/5 phosphorylation overlaps substantially in average transcription units.[43,44] On the other hand, interpretation is complicated by the fact that our pSer2/5 peptide has considerably more negative charge than the other peptides assayed and thus may be more prone to nonspecific effects. Interestingly, the pTyr1 peptide binds with a $K_d$ of 110 μM, which is tighter than that of the pSer2 (197 μM) and pSer5 (245 μM) peptides bearing an equivalent number of phosphate groups. A higher affinity for pTyr1 peptide with the same overall charge as pSer2 or pSer5 peptide is consistent with the similarity between nSH2 and well-characterized pTyr binding SH2 domains, but the physiological relevance of this result is not clear given the lack of observed pTyr modification of the RNAPII CTD in yeast.

The conclusion that the Spt6 tSH2 domain interacts specifically with phospho-CTD peptides is reinforced by our observations that an unphosphorylated form of the RNAPII CTD showed negligible ($K_d$ of >1000 μM) binding and that similar results were obtained when the assays were performed under a phosphate-buffered saline condition (data not shown). As a further test of specific interactions, we measured binding to a mutant form of Spt6(1247–1451) in which residues R1282 and S1284, which are important for phosphate binding in typical SH2 domains, were both substituted with




(c)

| CTD peptide | $K_D$ (μM)† WT | R1282A/S1284A |
|---|---|---|
| ○ pSer(2,5) | 23.0±2.2 | 90.9±14.2 |
| ● pTyr1 | 110.2±9.1 | 261.6±32.8 |
| × pSer2 | 197.1±39.69 | 228.4±17.8 |
| ▽ pSer5 | 245.4±29.4 | 348.1±56.7 |
| ✳ Unphos | >1000 | >1000 |

†Average $K_D$ ± standard error for at least 3 independent measurements

**Fig. 8.** Spt6 tSH2 binds RNAPII CTD phosphopeptides. (a) Peptides used in binding studies with positions of pSer and pTyr residues indicated. (b) Representative FA binding isotherms for Spt6(1247–1451) binding to various peptides with symbols defined in (c). (c) Binding affinities for WT and R1282A/S1284A Spt6(1247–1451) proteins based on FA experiments.

alanine. Binding to the pSer(2,5) peptide was decreased by ~4-fold, while binding to the pTyr1, pSer2, and pSer5 peptides was decreased by 2.4-fold, 1.2-fold, and 1.4-fold, respectively. These modest effects are consistent with the putative nSH2 phosphate binding site contributing to the binding interaction but not performing a dominant role for interaction with the peptides assayed.

Deletion of the entire tSH2 region by truncation of the *SPT6* gene leads to defects in growth attributed

to suboptimal transcription elongation.[22,23] To examine the importance of tSH2 residues implicated in pTyr binding *in vivo* more carefully, we mutated the single genomic copy of *SPT6* to produce proteins with R1282H, S1284D, R1286A, Q1303E, EN1313/1314AA, or K1343E mutation (nSH2 domain) or P1390A or K1411E mutation (cSH2 domain). The effects of these mutations were quite mild, failing to recapitulate the severe defects caused by truncation of the gene (data not shown; see Materials and Methods for a list of phenotypes screened). The C-terminal region of Spt6 therefore appears to have some activity that is not interrupted when residues within the tSH2 domain expected to be important for binding phosphorylated substrates are mutated.

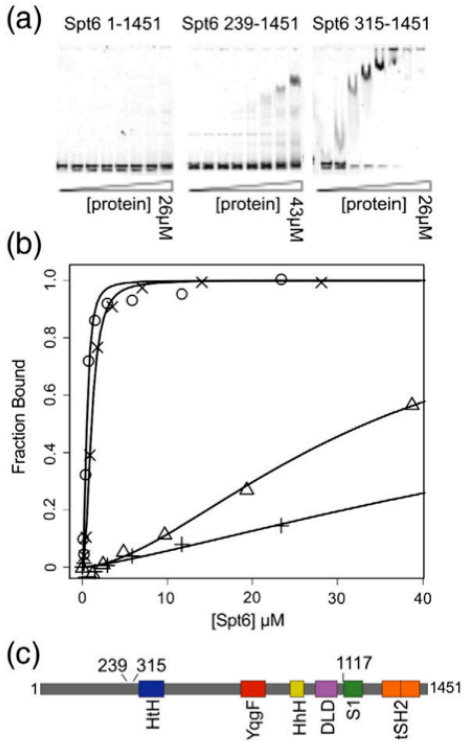



**Fig. 9.** dsDNA binding studies. (a) Representative gel shift assays for three different constructs of Spt6. (b) Representative binding isotherms used to calculate dissociation constants for various Spt6 constructs binding to the 177-bp Widom 601 dsDNA.[45] Symbols used to indicate isotherms for different constructs are as follows: (+) Spt6 (1–1451), $K_d = 106.7 \pm 38.9$ μM; (Δ) Spt6(239–1451), $K_d = 33.7 \pm 3.8$ μM; (×) Spt6(315–1117), $K_d = 1.08 \pm 0.06$ μM; (○) Spt6(315–1451), $K_d = 0.53 \pm 0.07$ μM. (c) Schematic diagram indicating endpoints for constructs used in DNA binding studies.

Our data are consistent with a recent report that concluded that pSer2 RNAPII CTD peptides bound the tSH2 domain of the *C. glabrata* Spt6 homolog with an affinity of 10 μM.[22] The tighter affinity observed in that study could reflect differences between the proteins but is more likely to be due to the very low (10 mM) concentration of NaCl used in the binding assays. Our data also extend the earlier work by showing that the *S. cerevisiae* Spt6 tSH2 domain displays little discrimination for binding CTD peptides with different modifications but does show a small preference for a peptide with a single phosphorylated tyrosine (pTyr1). Our findings suggest that Spt6 activities *in vivo* may be modulated by phosphorylation of binding partners and that the ligands of the tSH2 domain may include phosphorylated tyrosine residues.

### Binding to dsDNA

To test whether Spt6 is capable of binding dsDNA, we performed electrophoretic mobility gel shift assays using a 177-bp dsDNA fragment. We tested several different Spt6 constructs and found that binding was tighter ($K_d \pm$standard deviation) for Spt6(315–1451) ($K_d$ of $0.53 \pm 0.07$ μM) than for Spt6 (239–1451) ($K_d$ of $33.7 \pm 3.8$ μM) or Spt6(1–1451) ($K_d$ of $106.7 \pm 38.9$ μM) (Fig. 9), demonstrating that the disordered and negatively charged N-terminal residues diminish dsDNA binding. Unlike Tex, which requires the S1 domain for nucleic acid binding,[21] an Spt6 construct lacking the S1 domain, Spt6(315–1117), retains the ability to bind dsDNA with a $K_d$ of $1.08 \pm 0.06$ μM (Fig. 9b).

### Overall structure and functional implications

Our composite model of Spt6 (Fig. 10) features a core region (residues ~298–1117) that has multiple recognizable domains whose packing in the crystal likely reflects, to a large extent, their organization in solution, at least in the absence of binding partners. The N-terminal residues 1–297 display considerable overall negative charge and are expected to be highly mobile, while residues 239–263 also comprise the Spn1/Iws1-binding determinant and overlap with the nucleosome binding site.[14] This high degree of mobility may provide a flexible tether for bridging binding partners, such as Spn1/IwsI, RNAPII, and nucleosomes,[9] and the negative charge may modulate histone and DNA interactions.

Inherent flexibility is also a feature of the C-terminal S1 domain, H44, and tSH2 domain. The S1 domain is loosely associated with the core, lacks density in the Spt6(236–1259) structure, and may be visible in the Spt6(239–1451) structure only because of ordering by a crystal lattice contact. Whereas the core of the distantly related Tex protein clearly resembles that of Spt6, the Tex and Spt6 S1 domains are displaced

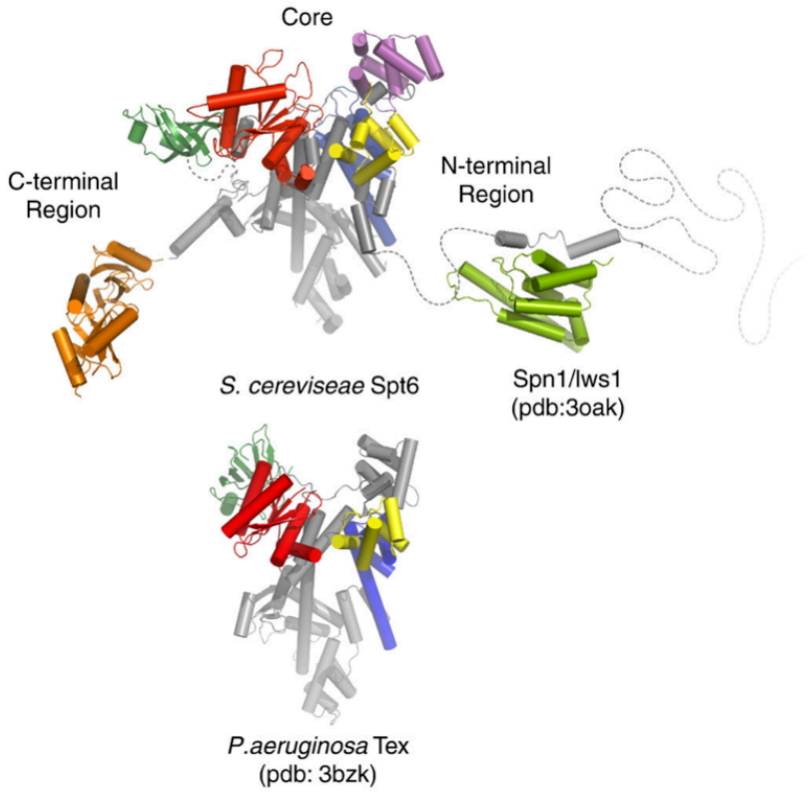



**Fig. 10.** Comparison of Spt6 and Tex structures. The overall structure of the Spt6 core resembles that of the prokaryotic Tex protein. This similarity implicates the Spt6 core in nucleosome-independent functions, such as transcriptional elongation on naked template DNA. The N-terminal region includes determinants essential for binding Spn1/Iws1 or nucleosomes, which appear to bind competitively with each other.[14] The negatively charged N-terminal region may also be important for modulating binding to nucleic acids. The C-terminal region has been implicated in binding RNAPII.[9]

by a rotation of ~80° and ~25- to 30-Å translation with respect to each other, and the S1 domain appears to be shifted by 14 Å in different crystal forms of Tex.[21] Therefore, it is likely that a highly mobile S1 domain is an important feature of both proteins, although one notable difference is that the Tex S1 binds nucleic acids, whereas our DNA binding data and consideration of surface amino acid residues indicate that the Spt6 S1 domain does not. Although the low conservation of residues in the putative binding cleft of the S1 domain suggests that this region is unlikely to have a highly conserved binding partner, the electrostatic characteristics of this surface (Fig. 5) are consistent with potential binding partners that are positively charged, such as histones.

H44 is visible only in the Spt6(236–1259) structure, where it also appears to be ordered by a lattice contact, and the C-terminal tSH2 domain is expected to be highly dynamic with respect to the rest of the protein. The leading model is that this domain binds RNAPII that is phosphorylated on Ser2 of its CTD.[9] Our binding data are consistent with this view, provided that other determinants contribute to binding/specificity, but also indicate the possibility that other binding partners might be functionally important. For example, Spt5 (discussed below) has been shown to co-localize with Spt6 and contains a phosphorylated C-terminal repeat domain similar to the RNAPII CTD,[46] which could be a physiological ligand for the Spt6 tSH2 domain.

Spt6 has been shown to be functionally associated with a large number of proteins involved in transcription elongation, chromatin maintenance, and RNA processing. The structure of Spt6 presented here will serve as a foundation for a more precise mapping of protein binding partners. Along with protein binding partners, Spt6 is also expected

to interact with nucleic acids in the transcription complex. DNA binding is probably important for nucleosome reassembly and potentially for transcription elongation. Association with nascent RNA transcripts could be important for enhancing the elongation rate or for organizing interactions with RNA modification/export factors such as REF/Aly. Future studies will be needed to see if Spt6 binds RNA and to further map nucleic acid binding to different domains of Spt6. For example, the YqgF or the $(HhH)_2$ domain may mediate binding to specific DNA structures such as four-way (Holliday) junction DNA, structures similar to those found at the DNA entry/exit points of nucleosomes.[47] The tSH2 domain may also contribute to nucleic acid interactions, as it is likely to bind negatively charged substrates containing phosphate groups. A simple electrostatic surface analysis indicates that each of these domains retains properties found in homologous domains with known functions. Furthermore, examination of the histone-binding activity of the various Spt6 domains will be of significant interest in furthering our understanding of nucleosome assembly/reassembly.

The structural similarity between Spt6 and the prokaryotic Tex protein is limited to the core and S1 domains. Consistent with the extent of structural similarity, N- and C-terminal regions that are unique to Spt6 are required for eukaryotic-specific interactions with nucleosomes,[11,13,14] hyperphosphorylated forms of the RNAPII CTD,[9] Spn1/IwsI,[14] and mRNA processing/export factors.[9] An attractive possibility is that the core region provides activities that are conserved among prokaryotes and eukaryotes. In this regard, it is striking that most of the Spt6 core domains belong to structural families whose members function in nucleic acid binding, an activity that is likely to be a key component of transcription factors such as Spt6 and Tex that are capable of stimulating elongation on nucleosome-free DNA templates.[9,16] Consistent with this possibility, our data indicate that the Spt6 core can bind dsDNA. Curiously, some of the putative nucleic-acid-binding surfaces of Spt6 domains are occluded in the structure, although conformational changes might displace residues 298–320 (H1 and H2) to expose a DNA-binding activity on the $(HhH)_2$ domain. Consistent with this model, we find that truncation of ~314 N-terminal residues leads to tighter dsDNA binding. One attractive model is that conformational changes of this nature are induced by binding partners such as histones or Spn1/Iws1. The relationship between Spt6 and Tex proposed here is reminiscent of that between another eukaryotic transcription elongation factor, Spt5, and its bacterial counterpart, NusG. These proteins also display similar core domains, while Spt5 has an acidic N-terminal extension and a C-terminal extension[48] that confers eukaryote-specific functions such as binding to RNAPII,[48] interaction with mRNA capping enzymes,[49] and extensive phosphorylation by RNAPII CTD kinases such as P-TEFb.[50] Therefore, like Spt5, Spt6 is likely to have built on the fundamental transcription activities of its core to accommodate the additional complexities of eukaryotic gene regulation.

## Materials and Methods

### Protein expression and purification

The protein constructs were expressed from pET151-D/TOPO vectors (Invitrogen) in BL21 codon plus (RIL) *E. coli* cells (Stratagene). Cultures were grown in autoinduction media[51] in baffled 1.8-l flasks at 37 °C with continuous shaking. After 4–8 h, the cultures were shifted to 23 °C and grown for an additional 16–24 h. Harvested cells were stored at −80 °C. Cells were thawed and lysed in buffer containing lysozyme and protease inhibitors, followed by sonication and centrifugation (25,000–30,000*g*). The soluble fraction was applied to nickel agarose resin (Qiagen) and eluted in buffer containing 300 mM imidazole and 100 mM NaCl, immediately followed by application to a heparin column (5-ml HiTrap Heparin; GE Healthcare Life Sciences) and elution over a NaCl gradient. Fractions containing Spt6 were pooled and processed overnight at room temperature in buffer containing tobacco etch virus protease. A nickel agarose column was used to remove the tagged tobacco etch virus protease and unprocessed Spt6 protein, and the flow-through was concentrated and loaded onto a size-exclusion column [Superdex 200 or S75 (for 1247–1451 constructs); GE Healthcare Life Sciences]. SeMet protein was expressed using an autoinduction protocol[51] for selenomethionine incorporation and purified by the same protocol as native protein. All crystals were grown in sitting drops, transferred briefly to a cryoprotection solution, suspended in a nylon loop, and plunged into liquid nitrogen (Table 1).

### Crystal structure determinations and refinements

Data were processed with HKL2000.[52] Spt6(236–1259) was determined by the multiple-wavelength anomalous diffraction method. SOLVE[53] was used to locate selenium atoms, and RESOLVE[54] was used for density modification and preliminary model building. AutoSol in PHENIX[26] was used to determine the Spt6(1247–1451) structure by the single-wavelength anomalous diffraction method. The Spt6 (239–1451) structure was determined by molecular replacement using AutoMR in PHENIX[26] to a resolution of 3.3 Å. A homology model built by TASSER[55] of the Spt6 S1 domain was used as a guide for model building. The native Spt6 (1247–1451) structure was determined by molecular replacement using Phaser.[56] PHENIX[26] and TLSMD[57] were used for refinement, Coot[58] was used for model building, and MolProbity[27] was used for structure validation. The following residues were ordered and included in the refined models: Spt6(236–1259) 297–455, 464–484, 501–561, 567–1002, 1009–1128, and 1219–1248; Spt6(239–1451) 312–455, 464–489, 509–552, 567–649, 653–1001, and 1014–1210; and

Spt6(1247–1451) 1247–1440. Structural alignments were performed using Dali[59] and SSM.[60] PyMOL[61] was used to create the figures. Electrostatic surface representations were calculated using PDB2PQR and APBS tools[62,63] using the AMBER force field and colored from red (−5 $kT/e$) to blue (+5 $kT/e$).

### DNA binding experiments

Widom 601 DNA (177 bp) with a 5′ Cy3 fluorophore was generated as described[45] followed by precipitation, gel purification, and electroelution. Electrophoretic mobility shift binding experiments were performed, and $K_d$ values were determined as described previously.[21] In short, 2-fold serial dilutions of the respective purified protein construct were mixed with nucleic acid substrate at room temperature in binding buffer [15 mM Tris–HCl (pH 7.5), 100 mM NaCl, 10% glycerol, and 0.5 mM ethylenediaminetetraacetic acid] where the final concentration of dsDNA was 10- to 20-fold below the estimated $K_d$ for the interaction. After incubation for 30 min, samples were run on 4–20% TBE native gels (Bio-Rad Laboratories) and imaged and quantified using a TYPHOON imaging system with ImageQuant software (GE Healthcare Life Sciences). The fraction bound was calculated by quantifying the $DNA_{total}$ (total fluorescence in entire lane) and $DNA_{free}$. DNA of slower mobility than the $DNA_{free}$ was considered bound. The fraction bound = $1-([DNA]_{free}/[DNA]_{total})$. Dissociation constants ($K_d$ values) were calculated by plotting data points and curve fitting in the program R[64] using the Hill formalism where fraction bound = $1/(1+(K_d n/[P]n))$. In all cases, standard deviations are calculated from at least three measurements, except for the Spt6(315–1117) construct, which was repeated twice.

### FA binding experiments

Peptides were synthesized by the University of Utah Core Facility or purchased commercially through AnaSpec Inc. (San Jose, CA), purified to >98% purity by HPLC, and confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Purified Spt6(1247–1451) was titrated in 1.5- to 2.0-fold serial dilutions against a constant concentration of fluorescein-labeled peptide (10- to 20-fold below estimated $K_d$) in 20 mM Tris–Cl (pH 7.5), 100 mM NaCl, and 5% glycerol. Samples were incubated at room temperature for at least 15 min prior to reading. Parallel and perpendicular fluorescence intensity was measured in a multi-well format using a Tecan Infinite 200 microplate reader using excitation/emission wavelengths of 485 nm/535 nm. Anisotropy values were calculated, normalized, and plotted as a function of protein concentration. $K_d$ values were determined by fitting the data using the equation[65] $A=((A_T\times([pro]/K_d))/(1+[pro]/K_d))$, where $A$ is the measured anisotropy, $A_T$ is the total change in anisotropy, and [pro] is the protein concentration.

### Genetic analysis of tSH2 mutations

The following alleles of *SPT6* were screened for phenotypes in strains isogenic with the A364a genetic background: wild type (WT), *spt6-R1282H*, *spt6-S1284D*, *spt6-R1286A*, *spt6-Q1303E*, *spt6-E1313A*, *N1314A*, *spt6-K1343E*, *spt6-P1390A*, and *spt6-K1411E*. Mutations were introduced into the genomic copy of *SPT6* such that expression was from the native promoter at the normal locus except for the introduction of a *URA3* or *TRP1* marker downstream of the open reading frame. Strains were tested for growth on rich medium at 30 °C and 38 °C; on medium lacking lysine at 30 °C and 37 °C (all strains had the *lys2-128∂* allele; thus, growth would reveal an $Spt^-$ phenotype); and on media containing 150 mM hydroxyurea, 75 μg/ml 6-azauracil, 0.6 μg/ml 4-nitroquinolone, 10 mM caffeine, 3% formamide, 1.2 M NaCl, 45 μg/ml mycophenolic acid, or 6% ethanol (all at 30 °C). None of the mutants were sensitive to any of the stress conditions relative to the WT strain. *spt6-S1284D* and *spt6-Q1303E* strains were somewhat more resistant to 3% formamide than the WT, and *spt6-Q1303E* strains displayed a very weak $Spt^-$ phenotype (faint growth after 7 days). Mutants were not tested for a defect in cryptic initiation.

### PDB accession numbers

Coordinates and structure factors for Spt6(236–1259), Spt6(239–1451), and SeMet Spt6(1247–1451) and native Spt6(1247–1451) have been deposited in the PDB with accession numbers 3psf, 3psi, 3psj, and 3psk, respectively.

## Acknowledgements

We thank Hua Xin and Charisse Kettelkamp for technical assistance and Heidi Schubert for advice with the crystallographic analysis. Portions of this work were performed in Core Facilities at the University of Utah, which were supported by P30CA042014 from the National Cancer Institute. Some of the X-ray diffraction data for this study were measured at the National Synchrotron Light Source (NSLS). Financial support for NSLS comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the U.S. Department of Energy and from the National Center for Research Resources of the National Institutes of Health (NIH). Portions of this research were performed at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and the NIH, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. S.J.J. was supported by a postdoctoral fellowship (GM074368). This work was supported by NIH grant RO1 GM076242.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2011.03.002

## References


1. Li, B., Carey, M. & Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, **128**, 707–719.
2. Luna, R., Gaillard, H., Gonzalez-Aguilera, C. & Aguilera, A. (2008). Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. *Chromosoma*, **117**, 319–331.
3. Clark-Adams, C. D. & Winston, F. (1987). The *SPT6* gene is essential for growth and is required for delta-mediated transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **7**, 679–686.
4. Kok, F. O., Oster, E., Mentzer, L., Hsieh, J. C., Henry, C. A. & Sirotkin, H. I. (2007). The role of the SPT6 chromatin remodeling factor in zebrafish embryogenesis. *Dev. Biol.* **307**, 214–226.
5. Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W. & Lis, J. T. (2009). Spt6 enhances the elongation rate of RNA polymerase II *in vivo*. *EMBO J.* **28**, 1067–1077.
6. Nishiwaki, K., Sano, T. & Miwa, J. (1993). *emb-5*, a gene required for the correct timing of gut precursor cell division during gastrulation in *Caenorhabditis elegans*, encodes a protein similar to the yeast nuclear protein SPT6. *Mol. Gen. Genet.* **239**, 313–322.
7. Shen, X., Xi, G., Radhakrishnan, Y. & Clemmons, D. R. (2009). Identification of novel SHPS-1-associated proteins and their roles in regulation of insulin-like growth factor-dependent responses in vascular smooth muscle cells. *Mol. Cell. Proteomics*, **8**, 1539–1551.
8. Baniahmad, C., Nawaz, Z., Baniahmad, A., Gleeson, M. A., Tsai, M. J. & O'Malley, B. W. (1995). Enhancement of human estrogen receptor activity by SPT6: a potential coactivator. *Mol. Endocrinol.* **9**, 34–43.
9. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev.* **21**, 160–174.
10. Vanti, M., Gallastegui, E., Respaldiza, I., Rodriguez-Gil, A., Gomez-Herreros, F., Jimeno-Gonzalez, S. *et al.* (2009). Yeast genetic analysis reveals the involvement of chromatin reassembly factors in repressing HIV-1 basal transcription. *PLoS Genet.* **5**, e1000339.
11. Adkins, M. W. & Tyler, J. K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Mol. Cell*, **21**, 405–416.
12. Kaplan, C. D., Laprade, L. & Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, **301**, 1096–1099.
13. Bortvin, A. & Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science*, **272**, 1473–1476.
14. McDonald, S. M., Close, D., Xin, H., Formosa, T. & Hill, C. P. (2010). Structure and biological importance of the Spn1–Spt6 interaction, and its regulatory role in nucleosome binding. *Mol. Cell*, **40**, 725–735.
15. Yoh, S. M., Lucas, J. S. & Jones, K. A. (2008). The Iws1: Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev.* **22**, 3422–3434.
16. Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D. K., Aida, M. *et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II *in vitro*. *Mol. Cell. Biol.* **24**, 3324–3336.
17. Andrulis, E. D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P. & Lis, J. T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature*, **420**, 837–841.
18. Kaplan, C. D., Holland, M. J. & Winston, F. (2005). Interaction between transcription elongation factors and mRNA 3′-end formation at the *Saccharomyces cerevisiae* GAL10–GAL7 locus. *J. Biol. Chem.* **280**, 913–922.
19. Pawson, T. (2004). Specificity in signal transduction: from phosphotyrosine–SH2 domain interactions to complex cellular systems. *Cell*, **116**, 191–203.
20. Maclennan, A. J. & Shaw, G. (1993). A yeast SH2 domain. *Trends Biochem. Sci.* **18**, 464–465.
21. Johnson, S. J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S. L. & Hill, C. P. (2008). Crystal structure and RNA binding of the Tex protein from *Pseudomonas aeruginosa*. *J. Mol. Biol.* **377**, 1460–1473.
22. Sun, M., Lariviere, L., Dengl, S., Mayer, A. & Cramer, P. (2010). A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II CTD. *J. Biol. Chem.* **285**, 41597–41603.
23. Diebold, M. L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J. & Romier, C. (2010). A non-canonical tandem SH2 enables interaction of elongation factor SPT6 with RNA polymerase II. *J. Biol. Chem.* **285**, 38389–38398.
24. Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
25. Ponting, C. P. (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* **30**, 3643–3652.
26. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J. *et al.* (2008). Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435.
27. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X. *et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383.
28. Dengl, S., Mayer, A., Sun, M. & Cramer, P. (2009). Structure and *in vivo* requirement of the yeast Spt6 SH2 domain. *J. Mol. Biol.* **389**, 211–225.
29. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635–645.
30. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. (2005). The many faces of the helix–turn–helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262.
31. Grant, R. P., Marshall, N. J., Yang, J. C., Fasken, M. B., Kelly, S. M., Harreman, M. T. *et al.* (2008). Structure of the N-terminal Mlp1-binding domain of the

*Saccharomyces cerevisiae* mRNA-binding protein, Nab2. *J. Mol. Biol.* **376**, 1048–1059.
32. Saito, A., Iwasaki, H., Ariyoshi, M., Morikawa, K. & Shinagawa, H. (1995). Identification of four acidic amino acids that constitute the catalytic center of the RuvC Holliday junction resolvase. *Proc. Natl Acad. Sci. USA*, **92**, 7470–7474.
33. Shao, X. & Grishin, N. V. (2000). Common fold in helix–hairpin–helix proteins. *Nucleic Acids Res.* **28**, 2643–2650.
34. Qiao, F. & Bowie, J. U. (2005). The many faces of SAM. *Sci. STKE*, **2005**, re7.
35. Park, H. H., Lo, Y. C., Lin, S. C., Wang, L., Yang, J. K. & Wu, H. (2007). The death domain superfamily in intracellular signaling of apoptosis and inflammation. *Annu. Rev. Immunol.* **25**, 561–586.
36. Theobald, D. L., Mitton-Fry, R. M. & Wuttke, D. S. (2003). Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 115–133.
37. Yu, E. Y., Wang, F., Lei, M. & Lue, N. F. (2008). A proposed OB-fold with a protein-interaction surface in *Candida albicans* telomerase protein Est3. *Nat. Struct. Mol. Biol.* **15**, 985–989.
38. Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D. & Kuriyan, J. (1993). Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell*, **72**, 779–790.
39. Lee, C. H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E. & Kuriyan, J. (1994). Crystal structures of peptide complexes of the amino-terminal SH2 domain of the Syp tyrosine phosphatase. *Structure*, **2**, 423–438.
40. Duyster, J., Baskaran, R. & Wang, J. Y. (1995). Src homology 2 domain as a specificity determinant in the c-Abl-mediated tyrosine phosphorylation of the RNA polymerase II carboxyl-terminal repeated domain. *Proc. Natl Acad. Sci. USA*, **92**, 1555–1559.
41. Lunde, B. M., Reichow, S. L., Kim, M., Suh, H., Leeper, T. C., Yang, F. *et al.* (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **17**, 1195–1201.
42. Ladbury, J. E., Lemmon, M. A., Zhou, M., Green, J., Botfield, M. C. & Schlessinger, J. (1995). Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. *Proc. Natl Acad. Sci. USA*, **92**, 3199–3203.
43. Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J. & Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278.
44. Phatnani, H. P. & Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936.
45. Lowary, P. T. & Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42.
46. Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W. H. *et al.* (2009). Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol. Cell. Biol.* **29**, 4852–4863.
47. Zlatanova, J. & van Holde, K. (1998). Binding to four-way junction DNA: a common property of architectural proteins? *FASEB J.* **12**, 421–431.
48. Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G. A. *et al.* (2008). Core structure of the yeast spt4–spt5 complex: a conserved module for regulation of transcription elongation. *Structure*, **16**, 1649–1658.
49. Pei, Y. & Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* **277**, 19639–19648.
50. Ivanov, D., Kwak, Y. T., Guo, J. & Gaynor, R. B. (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol. Cell. Biol.* **20**, 2970–2983.
51. Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif.* **41**, 207–234.
52. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In (Simon, M. I., Abelson, J. N., Carter, C. W. & Sweet, R. M., eds), Academic Press, New York, NY.
53. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **55**, 849–861.
54. Terwilliger, T. C. (2003). Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **59**, 38–44.
55. Roy, A., Kucukural, A. & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738.
56. McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **63**, 32–41.
57. Painter, J. & Merritt, E. A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **62**, 439–450.
58. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **60**, 2126–2132.
59. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 2780–2781.
60. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **60**, 2256–2268.
61. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. Delano Scientific, Palo Alto, CA.
62. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667.
63. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
64. Team, R. D. C.. (2010). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.
65. LiCata, V. J. & Wowor, A. J. (2008). Applications of fluorescence anisotropy to the study of protein–DNA interactions. *Methods Cell Biol.* **84**, 243–262.

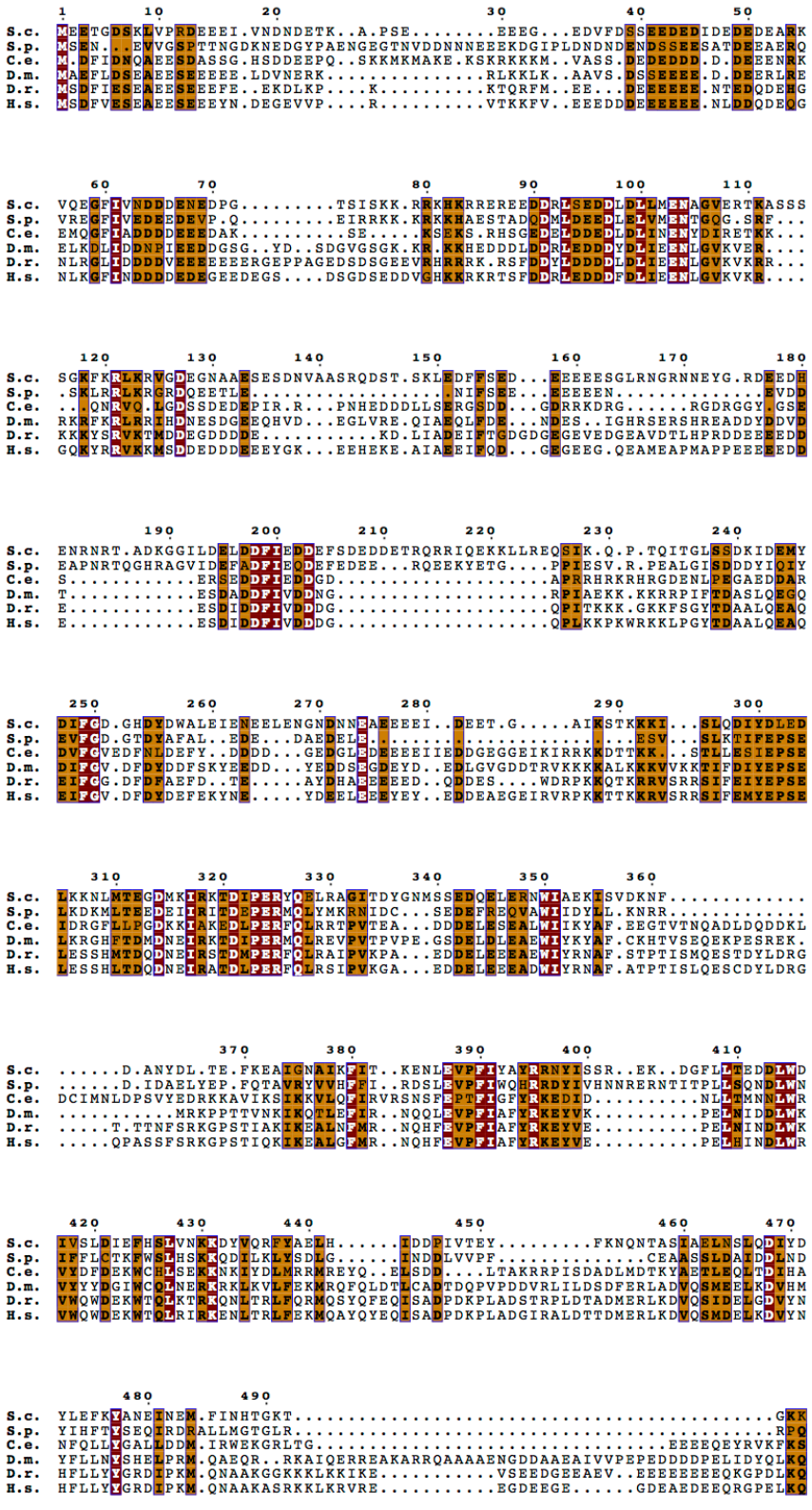
Supplemental Figure Legend

Figure S1. Spt6 Conservation.

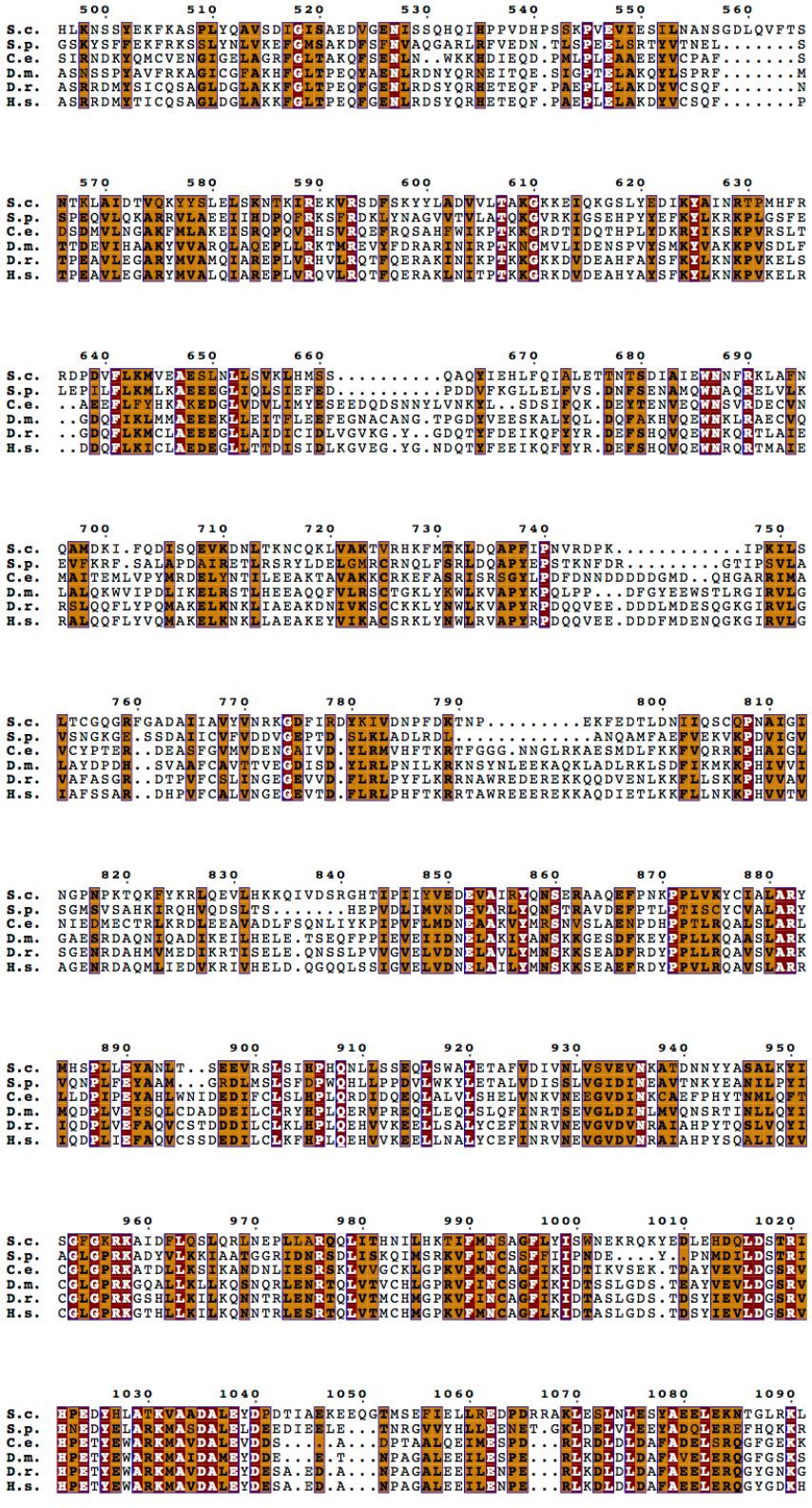
Spt6 protein from various eukaryotic species were aligned using T-Coffee[1] and ESPript[2] was used for visualization. Coloring of sequence represents degree of conservation, dark red background (invariant), orange font (conserved), in an alignment of proteins from *Saccharomyces cerevisiae* (S.c.), *Schizosaccharomyces pombe* (S.p.), *Caenorhabditis elegans* (C.e.), *Drosophila melanogaster* (D.m.), *Danio rerio* (D.r.), and *Homo sapiens* (H.s.).

1. Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-17.
2. Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research* **31**, 3320-3323.

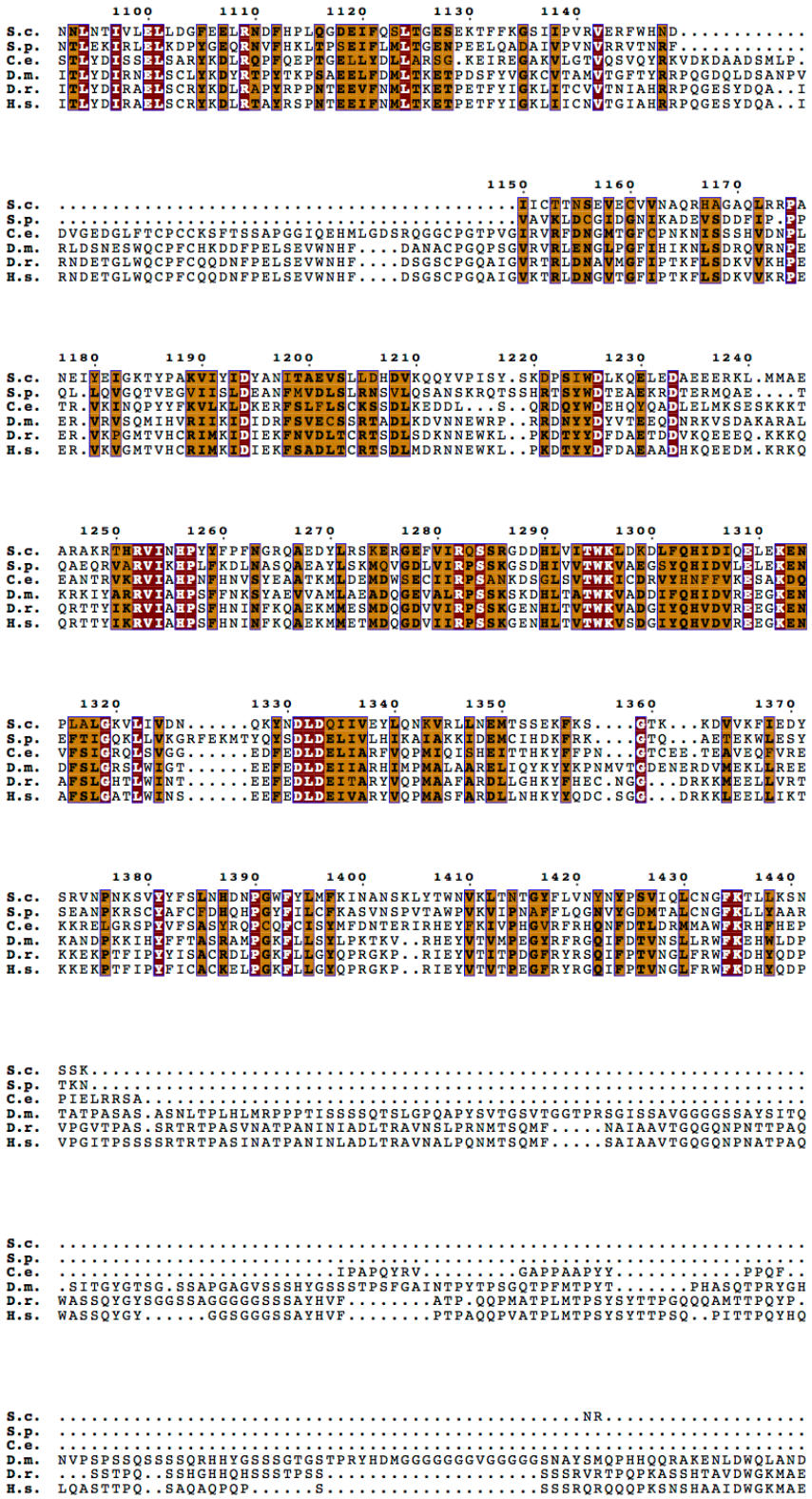
Supplemental Figure 1 Spt6 Crystal Structures Close, D et al



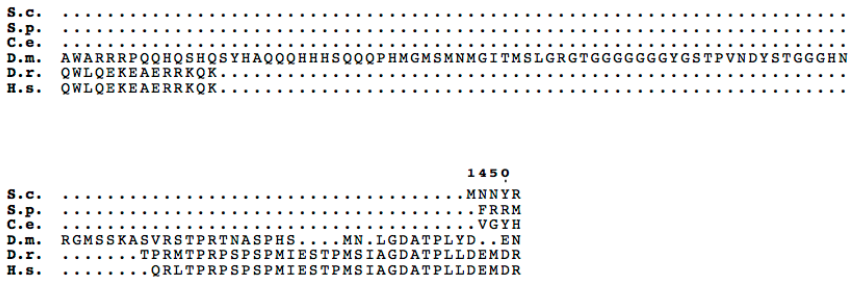
Supplemental Figure 1 Spt6 Crystal Structures Close, D et al



Supplemental Figure 1 Spt6 Crystal Structures Close, D et al



Supplemental Figure 1 Spt6 Crystal Structures Close, D et al



# CHAPTER 3

# MECHANISM OF PHOSPHORYLATION-DEPENDENT BINDING BETWEEN SPT6 AND RNAPII


## Summary

Spt6 is a conserved factor with roles in several facets of gene expression including transcription elongation and nucleosome reassembly following RNAPII passage. The C-terminus of Spt6 contains a tandem SH2 domain that comprises the only known SH2 domains in yeast. The Spt6 tSH2 directly binds RNAPII and is necessary for maximum recruitment of Spt6 to genes. Previous studies suggest that Spt6 tSH2 binds the phosphorylated RNAPII CTD, but this model did not stand up to rigorous biochemical scrutiny. We show that the Rpb1 linker, but not the CTD, is necessary and sufficient for binding the Spt6 tSH2. This interaction is enhanced by phosphorylation of Rpb1 S1493, and is ~500 fold tighter than the interaction with CTD-derived peptides. A crystal structure of the Spt6 tSH2-Rpb1 linker complex reveals a novel Spt6 phosphate binding pocket that coordinates phosphorylated S1493 (pS1493). Mutations that destabilize the interface cause loss of transcriptional repression in *S. cerevisiae*. These results suggest that Rpb1 S1493 phosphorylation recruits Spt6 to elongating complexes where it functions to reassemble nucleosomes immediately following RNAPII passage. Identification of the Rpb1 linker as the authentic Spt6 tSH2 binding site

requires a shift in the current model and re-evaluation of published data.

## Introduction

RNAPII is the twelve-subunit complex that transcribes all mRNAs and many noncoding RNAs. The largest subunit, Rpb1, contains a flexible C-terminal domain (CTD) consisting of heptad repeats of the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$. Phosphorylation of the CTD on tyrosine, serine, and threonine residues provides a dynamic binding surface that recruits chromatin regulators and mRNA processing factors at specific stages of transcription (Eick and Geyer, 2013). The CTD is tethered to the RNAPII core through a flexible ~80 residue linker.

Spt6 is a conserved transcription factor that co-localizes with RNAPII at sites of transcription (Kim et al., 2004; Mayer et al., 2010; Perales et al., 2013). It is essential in budding yeast and has multiple roles in gene expression including transcription, mRNA processing and export, histone posttranslational modification, and nucleosome positioning (Ardehali et al., 2009; Bucheli and Buratowski, 2005; Clark-Adams and Winston, 1987; Compagnone-Post and Osley, 1996; DeGennaro et al., 2013; Dronamraju and Strahl, 2014; Endoh et al., 2004; Hartzog et al., 1998; Ivanovska et al., 2011; Kaplan et al., 2005; Perales et al., 2013; Yoh et al., 2007; Yoh et al., 2008). Spt6 reassembles nucleosomes in the wake of elongating RNAPII, thereby re-establishing a repressive chromatin state that prevents aberrant transcription initiation (Adkins and Tyler, 2006; DeGennaro et al., 2013; Hainer et al., 2011; Ivanovska et al., 2011; Kaplan et al., 2003; Thebault et al., 2011). This activity is likely facilitated by direct interactions

with DNA, histones, and nucleosomes (Bortvin and Winston, 1996; Close et al., 2011; McCullough et al., 2015; McDonald et al., 2010). In addition to chaperoning histones, Spt6 further influences chromatin state by promoting histone modifications that are associated with active transcription (DeGennaro et al., 2013; Yoh et al., 2008). While altering chromatin state indirectly affects RNAPII elongation rates, Spt6 also stimulates RNAPII elongation on chromatin free DNA *in vitro* (Endoh et al., 2004; Yoh et al., 2007) and on permissive chromatin *in vivo* (Ardehali et al., 2009). These results establish Spt6 as a *bona fide* transcription elongation factor.

In addition to its roles in nucleosome assembly and transcription elongation, Spt6 is important for mRNA processing and export. In yeast, Spt6 mutants are defective in transcription termination, 3' processing, and mRNA export (Bucheli and Buratowski, 2005; Estruch et al., 2009; Kaplan et al., 2005). Consistent with these observations, Spt6 genetically interacts with the mRNA export factors Mex67 and Yra1 (Burckin et al., 2005; Estruch et al., 2009). In mammals, Spt6 is required for proper splicing of an HIV transcript and nuclear export of mRNA (Yoh et al., 2007). Furthermore, Spt6 physically associates with the RNA processing exosome in *Drosophila* (Andrulis et al., 2002).

Spt6 comprises 1,451 residues that can be roughly divided into three structural regions (Close et al., 2011). The N-terminal ~300 residues are highly acidic, predicted to be disordered, and necessary for binding both nucleosomes and the transcription factor Spn1 (Diebold et al., 2010a; McDonald et al., 2010). The core of Spt6 contains several domains with structural similarity to known

DNA and protein binding domains and has overall similarity to the prokaryotic transcription factor *Tex* (Johnson et al., 2008). The C-terminal ~200 residues comprise a tSH2 domain containing the only known SH2 domains in yeast (Close et al., 2011; Diebold et al., 2010b; Sun et al., 2010). The nSH2 domain resembles a canonical phosphotyrosine binding module, but the cSH2 domain appears cryptic.

The Spt6 tSH2 domain is important for Spt6 function *in vivo*. In yeast, deletion of the tSH2 domain causes slow growth and phenotypes attributed to defects in transcription elongation (Diebold et al., 2010b; Hartzog et al., 1998; McCullough et al., 2015; Sun et al., 2010). In mammals, the tSH2 domain is important for mRNA processing and export, and for class switch recombination (Begum et al., 2012; Okazaki et al., 2011; Yoh et al., 2007). Furthermore, deletion of the tSH2 domain reduces recruitment of Spt6 throughout transcribed genes (Burugula et al., 2014; Mayer et al., 2010). At least one activity of the Spt6 tSH2 domain is direct binding to hyperphosphorylated RNAPII (Yoh et al., 2007).

The prevailing interpretation of the data is that Spt6 tSH2 binds RNAPII through the phosphorylated RNAPII CTD. While an attractive model, thorough biochemical analysis from several labs using different techniques does not support a biologically relevant interaction (Close et al., 2011; Diebold et al., 2010b; Liu et al., 2011; Mayer et al., 2012; Sun et al., 2010). The affinities of Spt6 tSH2 for phosphorylated CTD-derived peptides were relatively weak *in vitro* (~1 μM to 5 mM depending on salt concentration, technique, and peptide), and there was little preference for different phosphorylation states, suggesting

nonspecific interactions. Consistent with a nonspecific interaction, Spt6 tSH2 mutations expected to greatly alter binding to a phosphorylated peptide had only modest effects. Collectively, these observations suggest an alternative mechanism for Spt6 binding to RNAPII.

We show that Spt6 tSH2 directly binds the flexible, ~80 residue Rpb1 linker that tethers the CTD to the RNAPII core. Importantly, this interaction requires phosphorylation of the Rpb1 linker on a specific, conserved serine residue. Using X-ray crystallography and biochemistry, we show that a novel phosphate-binding pocket in the cryptic Spt6 cSH2 domain coordinates the phosphoserine. Mutating the Spt6-Rpb1 linker interface in yeast causes the Spt$^-$ phenotype and cryptic promoter activation, suggesting defects in maintaining repressive chromatin. These data support a model in which Rpb1 linker phosphorylation recruits Spt6 to elongation complexes where it is positioned to reassemble nucleosomes immediately following RNAPII passage.

## Results

### The Rpb1 Linker Is Necessary for Spt6 tSH2 Binding

Spt6 tSH2 is thought to bind RNAPII through the CTD, but it has never been shown that the CTD is necessary for this interaction. The RNAPII CTD is required for viability, but insertion of protease sites adjacent to the CTD is tolerated (Li and Kornberg, 1994). We inserted PreScission Protease sites into Rpb1 that facilitated removal of the CTD (ΔCTD) or linker and CTD (Δlinker) during purification (Figures 3-1A and 3-1B). Purified RNAPII proteins were transferred to a nitrocellulose membrane and tested for binding to purified

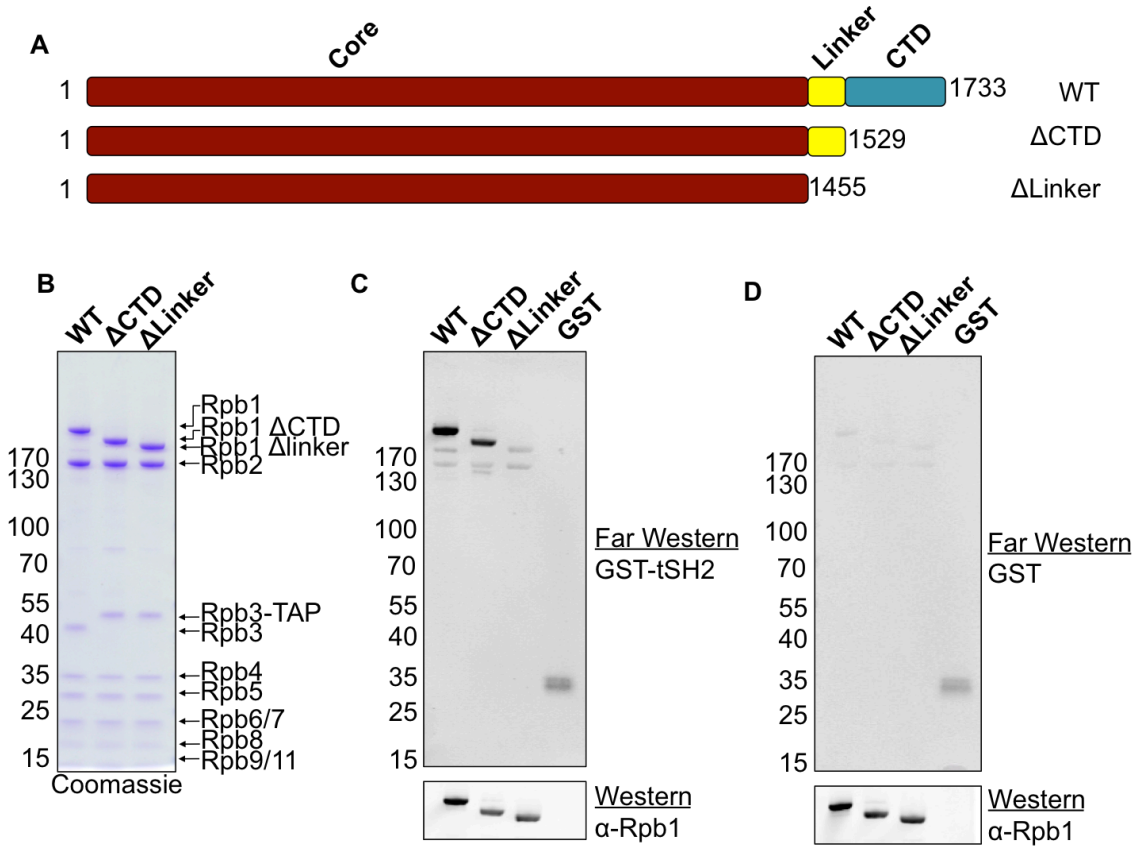



**Figure 3.1. The Rpb1 linker is necessary for binding Spt6 tSH2.**
A) Schematic representation of full-length Rpb1 (WT) and deletion constructs tested for binding (ΔCTD and ΔLinker).
B) Coomassie stained gel of purified RNAPII containing the Rpb1 constructs depicted in A. The Rpb3 subunit of ΔCTD and ΔLinker RNAPII has a residual calmodulin tag, causing slower migration compared to wild type.
C) Far western blot using GST-Spt6$^{1223\text{-}1451}$ to probe a membrane containing the purified proteins shown in B. A western blot using an antibody that recognizes the N-terminus of Rpb1 was used to determine transfer efficiency.
D) As in C, but the membrane was probed with free GST.

recombinant GST-Spt6 tSH2 (GST-Spt6$^{1223-1451}$) in a far western blot. Consistent with previous observations (Yoh et al., 2007), GST-Spt6$^{1223-1451}$ bound full-length Rpb1 (residues 1-1733) (Figure 3.1C). Free GST does not bind RNAPII, indicating that the interaction is specific to the tSH2 domain (Figure 3.1D). Surprisingly, Spt6 tSH2 retained complete binding to Rpb1 ΔCTD (residues 1-1529). Removal of the 74 residue linker that tethers the CTD to the RNAPII core, however, abolished binding (Figure 3.1C).

A possible interpretation of these results is that Spt6 tSH2 can bind both the linker and CTD, such that binding is observed as long as at least one binding site is present. To address this, we identified residues that may contribute to binding by aligning Rpb1 linker sequences across eukaryotes. Just six residues that cluster into two groups of three are strictly conserved: Cluster 1) T1471, P1472, and Y1473 and Cluster 2) F1492, S1493, and P1494 (Figure 3.2A). We performed alanine scanning mutagenesis of these residues to see if they are important for binding in a far western blot (Figure 3.2B and C). Mutating residues in Cluster 1 modestly reduced binding with Rpb1$^{T1471A}$ having the largest effect. Mutating any of the residues in Cluster 2, on the other hand, greatly reduced binding. Bound Spt6 was reduced nearly to background in the Rpb1$^{S1493A}$ mutation and reduced 4.2 and 3.2 fold in Rpb1$^{F1492A}$ and Rpb1$^{P1494A}$, respectively (Figure 3.2C). These results show that three conserved residues in the Rpb1 linker are necessary for binding and indicate that the CTD is not sufficient for interaction with Spt6.

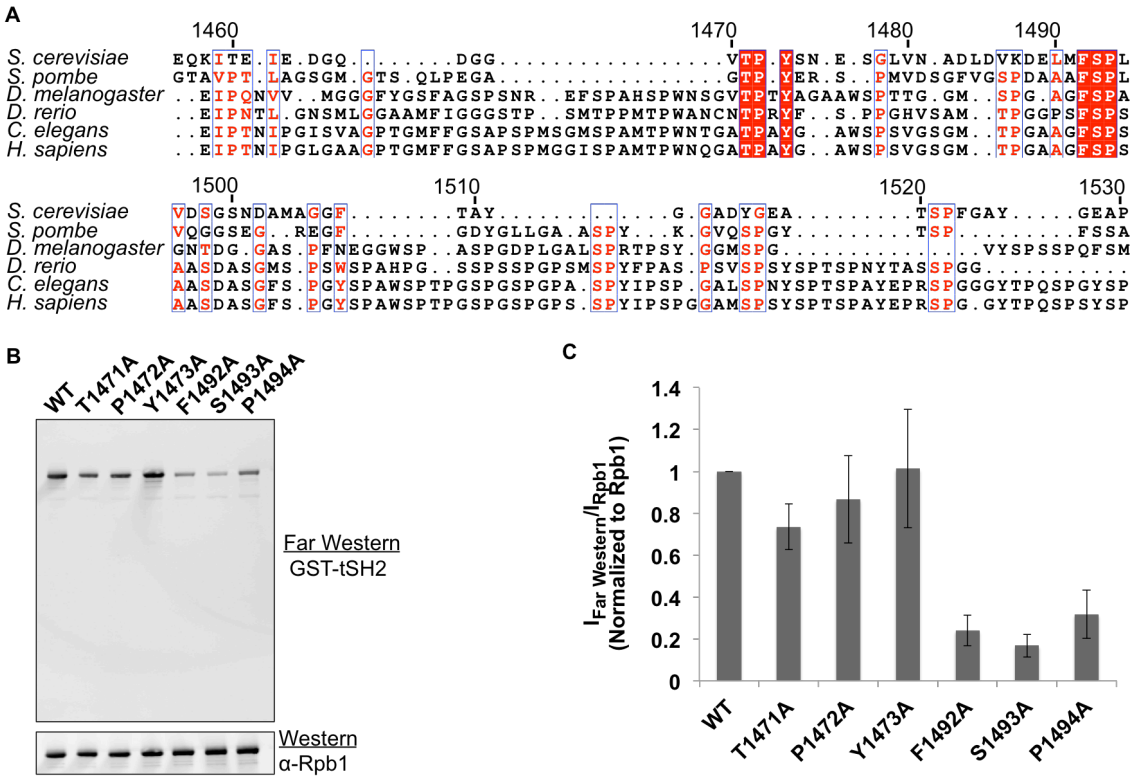



**Figure 3.2. Mutating conserved residues in the Rpb1 linker disrupt binding even when the CTD is intact.**

A) Alignment of the Rpb1 linker across eukaryotic species. Rpb1 sequences from the indicated organisms were aligned using T-Coffee (Notredame et al., 2000) and visualized using ESPript (Robert and Gouet, 2014). Residues in the linker are shown and numbered according to the *S. cerevisiae* sequence. White characters on a red background represent strictly conserved residues and red font on a white background represents highly similar residues (global similarity score $\geq$ 0.7).

B) Far western blots testing Rpb1 linker mutants for binding to GST-Spt6$^{1223\text{-}1451}$. Transfer efficiency was determined by blotting with an antibody that recognizes the Rpb1 CTD (8WG16: Covance)

C) Quantification of the far western blot. Ratios of the far western blot signal to the Rpb1 blot signal were calculated to corrected for differences in loading and transfer efficiency. Ratios were normalized to Rpb1. The averages from at least three independent experiments are shown. Error bars indicate the standard deviation.

Spt6 tSH2 Binds Rpb1 Phosphorylated on Residue S1493

To determine if the Rpb1 linker is sufficient for Spt6 tSH2 binding, we used fluorescence anisotropy to quantify binding to a synthesized linker peptide with an N-terminal fluorescein (Fl-Rpb1$^{1476-1511}$). This peptide includes residues F1492, S1493, and P1494. The peptide bound purified recombinant Spt6 tSH2 (Spt6$^{1247-1451}$) with a $K_D$ greater than 30 μM (Figure 3.2A), which is likely too weak to be detected in a far western blot. S1493 and P1494 form a minimal cyclin-dependent kinase phosphorylation site, prompting us to consider that binding could be enhanced by S1493 phosphorylation. We tested binding to Fl-Rpb1$^{1476-1511}$ with S1493 phosphorylated (Fl-Rpb1$^{1476-1500\ pS1493}$). The $K_D$ for the phosphorylated peptide was 70 nM, which is approximately 500 fold tighter than the unphosphorylated peptide (Figure 3.3A). Furthermore, the affinity is approximately 3 orders of magnitude tighter than phosphorylated CTD-derived peptides under the same buffer conditions (Close et al., 2011). Our assay was validated using a competitive fluorescence anisotropy experiment in which an unlabeled Rpb1 linker peptide was used to displace Spt6$^{1247-1451}$ from FL-Rpb1$^{1476-1511\ pS1493}$. The $K_I$ in this assay was 138 nM, which is in reasonably close agreement with the $K_D$ observed in direct biding assays (Figure 3.3B).

Rpb1 S1493 phosphorylation has not been reported in the literature. To determine if this modification exists *in vivo*, we raised an antibody against a phosphorylated linker peptide. The specificity of the antibody for the phosphorylated peptide versus the unphosphorylated peptide was validated by EMSA. The antibody was used in a western blot on yeast whole cell extracts

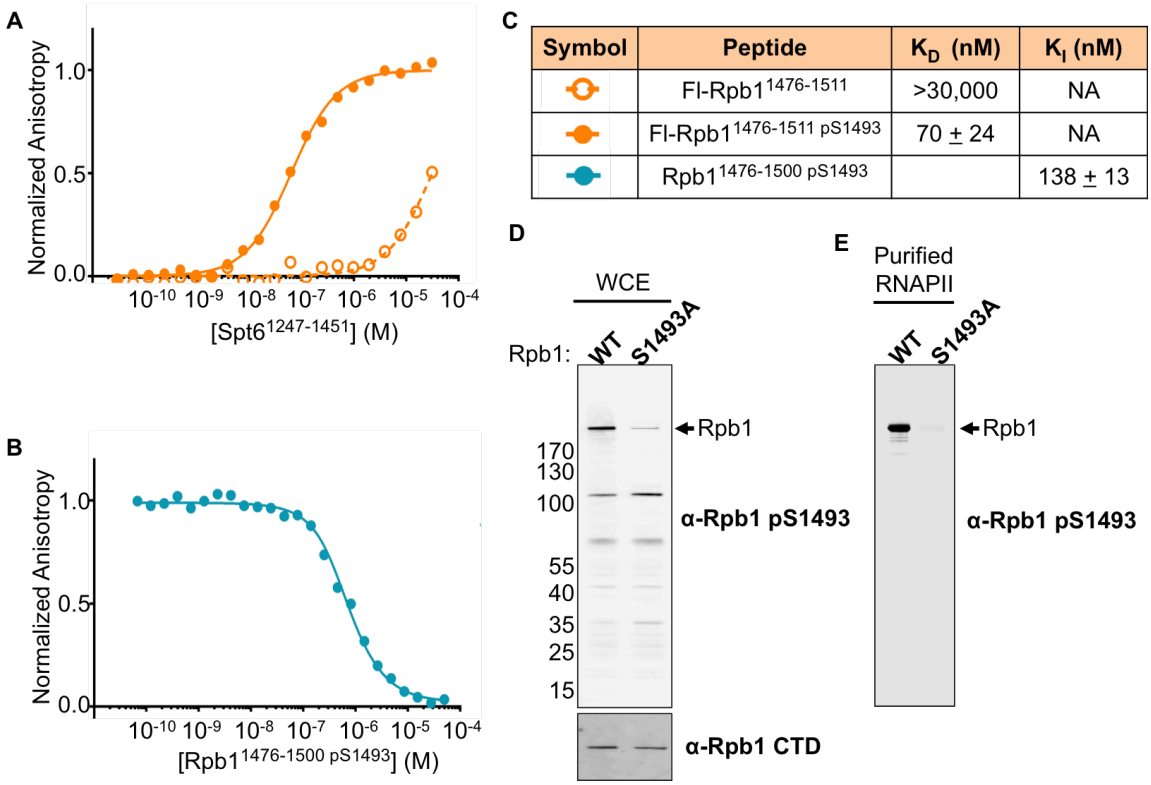



**Figure 3.3. Spt6 tSH2 binds the Rpb1 linker phosphorylated on S1493.**
A) Representative isotherms for Spt6 binding to Rpb1 linker peptides determined by fluorescence anisotropy. Spt6$^{1247\text{-}1451}$ was titrated into Fl-Rpb1$^{1476\text{-}1511}$ peptides with or without S1493 phosphorylated.
B) Competition fluorescence anisotropy assay. Unlabeled Rpb1$^{1476\text{-}1500}$ was titrated into constant concentrations of Spt6$^{1247\text{-}1451}$ (200 nM) and Fl-Rpb1$^{1476\text{-}1511\ pS1493}$ (0.5 nM).
C) Binding affinities of Spt6$^{1247\text{-}1451}$ for Rpb1 linker peptides determined by fluorescence anisotropy. Error estimates shown are for the standard deviation determined from at least three independent experiments.
D) Western blot on yeast whole cell extracts harboring wild type (WT) or mutant (S1493A) Rpb1 using an antibody raised against a phosphorylated peptide derived from the Rpb1 linker.
E) Western blot on purified RNAPII with WT or S1493A versions of Rpb1 using an antibody raised against a phosphorylated peptide derived from the Rpb1 linker.

resolved by SDS-PAGE (Figure 3.3D). Several bands were observed including a band that migrates at the same molecular weight as Rpb1. The intensity of this band is greatly diminished when S1493 is mutated to an alanine, confirming that the antibody preferentially recognizes the phosphorylated form of Rpb1. These observations were confirmed in a western blot with purified RNAPII (Figure 3.3E). Overall, these results demonstrate that Rpb1 S1493 is phosphorylated *in vivo* to promote binding to Spt6 tSH2.

Crystal Structure of the Spt6 tSH2-Rpb1 Linker Complex

To understand how Spt6 tSH2 binds the serine phosphorylated linker, we determined a crystal structure of Spt6 tSH2 bound to a phosphorylated Rpb1 peptide. The structure of $Spt6^{1247\text{-}1451}$ in complex with $Rpb1^{1476\text{-}1500\ pS1493}$ was determined by molecular replacement using the coordinates of yeast $Spt6^{1247\text{-}1451}$ (PDB: 3PSJ) (Close et al., 2011) and refined against 2.2 Å data to $R_{work}/R_{free}$ values of 17.7/24.3 (Figure 3.4 and Table 3.1). The crystals contained one complex in the asymmetric unit. Spt6 residues T1250-R1451 and Rpb1 residues L1479-S1498 were visible in the electron density map. Spt6 residues 1441-1451 were not ordered in previous crystal structures (Close et al., 2011), but are visible here due to lattice contacts. There is no density for the three N-terminal residues of Spt6 or the three N-terminal and two C-terminal residues of Rpb1, which are not included in the refined model. Binding to the Rpb1 linker did not induce conformational changes in Spt6 tSH2, which superimposes on unbound tSH2 (PDB ID: 3PSJ) with an RMSD of ~0.52 Å over 189 pairs of Cα atoms.

The Rpb1 linker interacts exclusively with the non-canonical cSH2

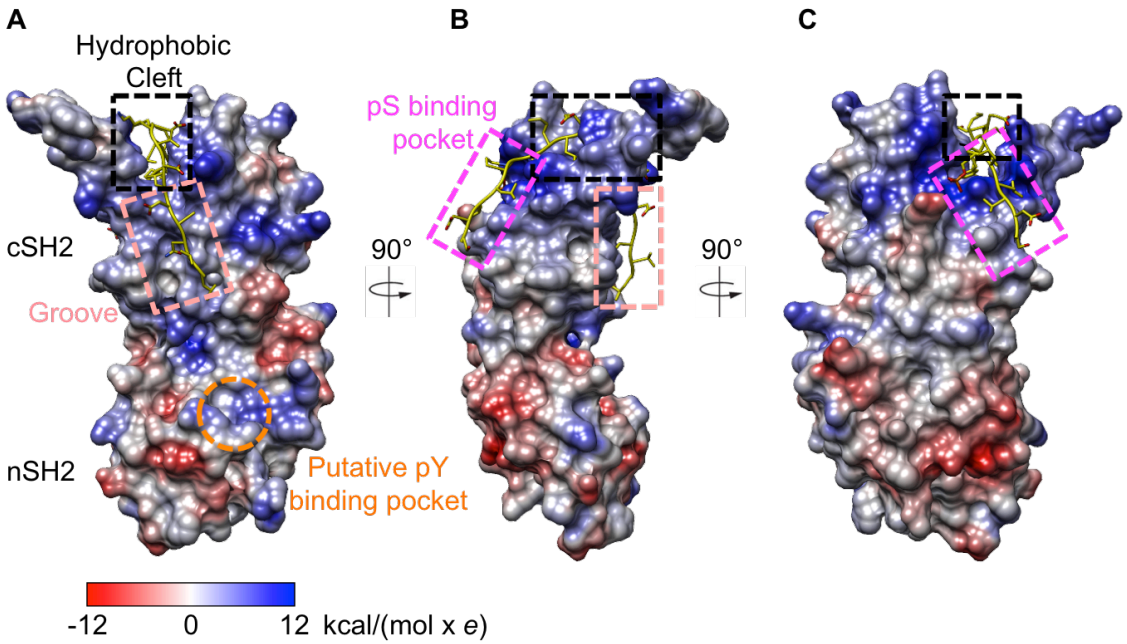



**Figure 3.4. Structure of the Spt6 tSH2-Rpb1 linker complex.**
A) Spt6 tSH2 is shown as a surface representation colored by electrostatic surface potential (±kcal/(mol x *e*)). The nSH2 and cSH2 domains are indicated. The Rpb1 linker peptide is colored yellow and shown as sticks. A dashed orange circle indicates the putative Spt6 cSH2 phosphotyrosine binding pocket. The groove and hydrophobic cleft are highlighted by dashed peach and black boxes respectively.
B) Orthogonal view of A. Colored as in A. The phosphoserine binding pocket is highlighted by a dashed magenta box.
C) Orthogonal view of B. Colored as in A and B.

Table 3.1

Crystallographic Data Collection and Refinement

| | |
|---|---|
| Data Collection | |
| Beamline | Home source |
| Wavelength | 1.54178 |
| Space group | $C222_1$ |
| Unit cell dimensions: a, b, c (Å) | 43.014, 103.54, 115.487 |
| Resolution | 27.6-2.20 (2.3-2.2) |
| Completeness (%) | 89.81 (60) |
| Redundancy | 24.7 (17.0) |
| Refinement | |
| Resolution | 27.6-2.20 (2.3-2.2) |
| Number of reflections | 12,178 |
| $R_{work}/R_{free}$ (%) | 17.7/24.3 |
| Number of protein atoms | 1,859 |
| Number of water molecules | 83 |
| Rmsd bond lengths (Å)/angles (°) | 0.007/0.878 |
| Ramachandran favored/allowed (%) | 96.33/3.67 |
| Values in parenthesis are for the highest resolution shell. | |

domain, draping over the top of Spt6 tSH2 in an extended conformation that buries 1,194 Å$^2$ of solvent accessible surface area. Extensive hydrophobic and polar contacts contribute to the interaction. The binding surface can be divided into three regions: a basic phosphoserine binding pocket, a hydrophobic cleft, and a shallow groove (Figure 3.4). The phosphoserine binding pocket is on the cSH2 domain on the tSH2 face opposite the putative phosphotyrosine binding pocket. It is formed by residues K1355, K1435, and Y1381 (Figures 3-4C and 3-5A). pS1493 nestles in this pocket and forms hydrogen bonds between the phosphate and the amino groups of K1355 and K1435 as well as the hydroxyl group of Y1381. The amino group of Spt6 K1435 is also in position to hydrogen bond with the main chain carbonyl of P1494. This extensive network of hydrogen bonds explains the importance of phosphorylation for binding. The binding surface continues C-terminal to pS1493. In this region, Rpb1 L1496 makes van der Waals contacts with the hydrocarbon side chain of Spt6 K1435, and the main chain nitrogen and carbonyl of Rpb1 V1497 hydrogen bonds with the Spt6 N1432 carboxamide. Additional interactions that are N-terminal to pS1493 are mediated through the Spt6 hydrophobic cleft and shallow groove.

The hydrophobic cleft formed by Spt6 residues Y1381, F1397, I1399, Y1406, W1408, F1434, and L1438 cuts through the top of Spt6 tSH2 (Figure 3.4). Rpb1 residues F1492, L1490, and L1484 bury 411 Å$^2$ of solvent accessible surface area in this cleft (Figure 3.5B and C). In addition to the hydrophobic interactions, several polar contacts contribute to binding in this region. The main chain carbonyls of Rpb1 M1491 and L1484 form hydrogen bonds with the main

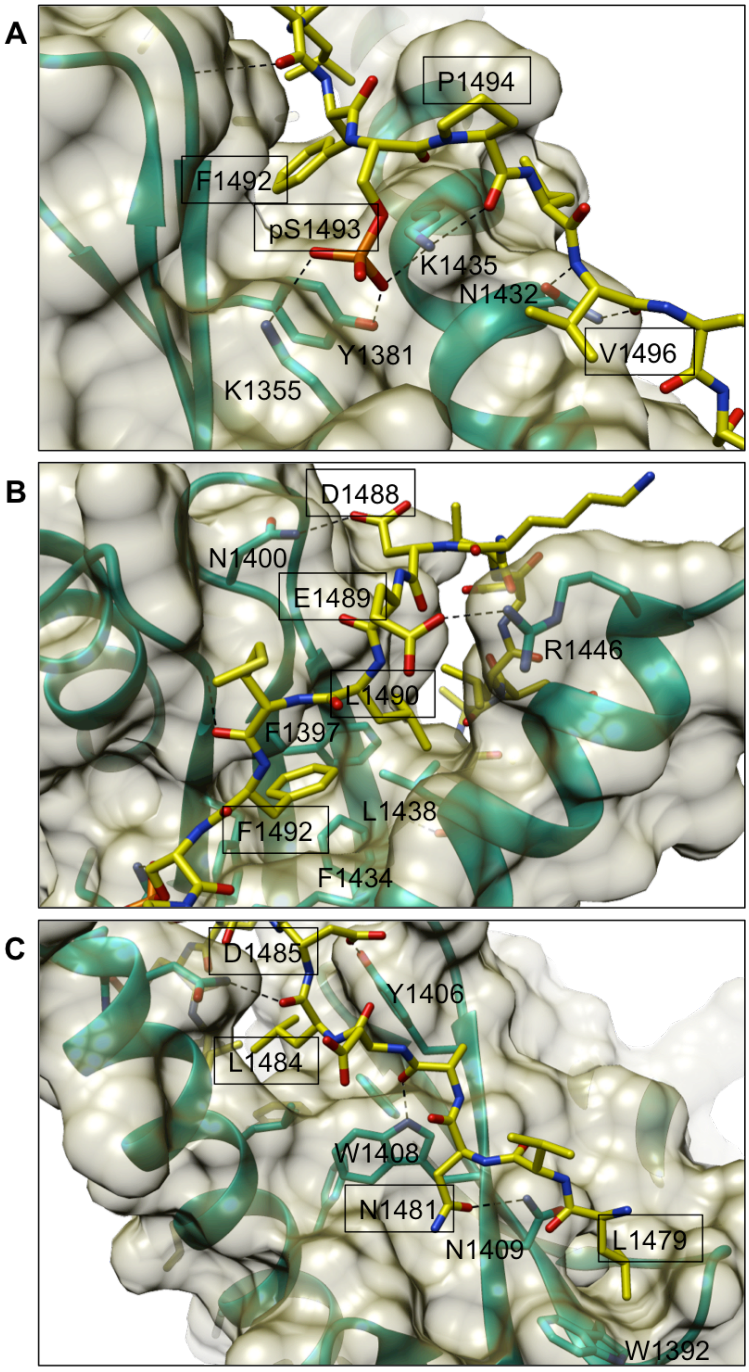



**Figure 3.5. Details of the Spt6-Rpb1 interface.**
A) Close up view of the pS1493 binding pocket. Spt6 is shown as a cyan ribbon with a semitransparent tan surface. Residues that contact Rpb1 are shown as sticks and labeled without boxes. Rpb1 is shown as yellow sticks. Residues that contact Spt6 are labeled in boxes with black borders. Hydrogen bonds are shown as black dashed lines.
B) Close up view of the hydrophobic binding cleft. Colored as in A.
C) Close up view of part of the hydrophobic binding cleft and the Rpb1 binding groove on the front of Spt6 tSH2. Colored as in A.

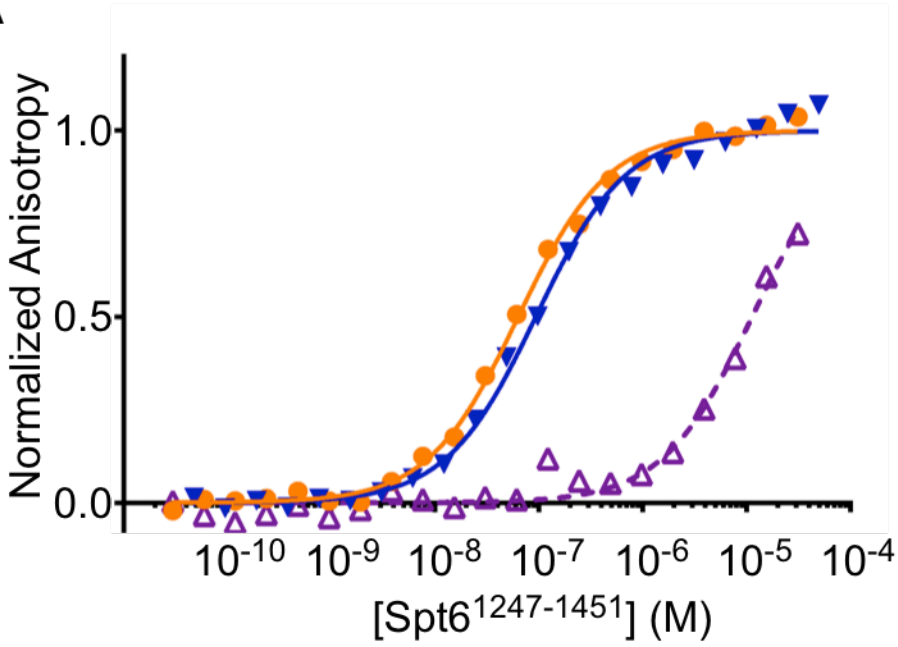
chain nitrogen of Spt6 S1379 and the carboxamide of Spt6 N1445, respectively. Additional hydrogen bonds are formed between the carboxylate groups of Rpb1 E1489, D1488, and D1485 and the hydroxyl, carboxamide, and guanidinium groups of Spt6 Y1406, N1400, and R1446, respectively.

The shallow binding groove runs down the front of Spt6 cSH2 and adds moderate contributions to the binding surface (Figures 3-4A and 3-5C). The carboxamides of Rpb1 N1481 and Spt6 N1409 hydrogen bond as do the main chain carbonyl of Rpb1 A1482 and the indole nitrogen of Spt6 W1408. At the N-terminus of Rpb1, L1479 contacts a hydrophobic patch formed by Spt6 W1392. Consistent with the groove contributing less to binding, the electron density for Rpb1 was not as strong in this region.

## Mutations in the Spt6 tSH2-Rpb1 Linker Interface Disrupt Binding *In Vitro*

Our structure explains the loss of Spt6 binding when Rpb1 F1492 is mutated or S1493 is not phosphorylated. To further validate the crystallographic interface, we measured binding affinities of purified Spt6$^{1247\text{-}1451}$ mutants to Rpb1$^{1476\text{-}1511\ pS1493}$ using fluorescence anisotropy (Figure 3.6). The Spt6 K1435A protein bound Fl-Rpb1$^{1476\text{-}1511\ pS1493}$ with a Kd greater than 10 μM, an ~150 fold reduction in affinity compared to wild type. Mutations in the canonical phosphotyrosine binding pocket (R1282A/S1284A), on the other hand, did not affect binding. These results validate the crystallographic interface and demonstrate the importance of phosphoserine binding pocket residues for binding *in vitro*.

A




B

| Symbol | Spt6$^{1247-1451}$ | $K_D$ (nM) |
|---|---|---|
| ● | WT | 70 ± 24 |
| ▼ | R1282A/S1284A | 93 |
| △ | K1435A | >10,000 |

**Figure 3.6. Mutations in the Spt6-Rpb1 interface disrupt binding *in vitro*.**
A) Representative Spt6-Rpb1 binding isotherms determined by fluorescence anisotropy. Spt6$^{1247-1451}$ wild type (orange circles) and mutant (R1282A/S1282A, closed blue triangles; K1435A, open purple triangles) proteins were titrated into Fl-Rpb1$^{1476-1511\ pS1493}$.
B) Binding affinities based on experiments in A. Error estimates shown are for the standard deviation determined from at least three independent experiments.

The Spt6 tSH2-Rpb1 Linker Interaction Is Important *In Vivo*

To determine the importance of Spt6 tSH2 binding to the Rpb1 linker *in vivo*, we introduced mutations that disrupt the interaction into the native *SPT6* and *RPB1* genes at the normal genomic loci. All of the mutations caused the Spt$^-$ phenotype, indicating loss of transcriptional repression due to defects in chromatin maintenance (Figure 3.7A). Importantly, the severity of the phenotype correlated with the severity of the mutation on binding. The *spt6-KK-AA* mutant replaces the two lysines in the phosphoserine binding pocket with alanines, and causes a moderate Spt$^-$ phenotype. This phenotype was recapitulated by *rpb1-S1493*, which eliminates phosphorylation by mutating Rpb1 S1493 to an alanine. The phenotypes of *spt6-KK-AA* and *rpb1-S1493* are partially additive, suggesting that each allele causes only a partial loss of chromatin maintenance. This could indicate that each mutation results in a partial loss of binding, or that these interfaces have additional functions. A more severe Rpb1 mutation (*rpb1-1490-1495MKRRRK*) that replaces residues L$^{1490}$MFSPLV$^{1496}$ with K$^{1490}$MKRRK$^{1496}$ causes a very strong Spt$^-$ phenotype. These yeast are able to grow on media lacking lysine almost as well as on complete medium, suggesting that the Spt$^-$ phenotype is saturated. It is therefore unsurprising that the Spt$^-$ phenotypes of *rpb1-1490-1495MKRRRK* and *spt6-KK-AA* are not additive (Figure 3.7A), which is consistent with the model that interaction through these Rpb1 residues mediates the entire Spt$^-$ phenotype of Spt6.

Interestingly, this *rpb1-1490-1495MKRRRK* strain also displays activation of a cryptic promoter reporter. Similar to the Spt$^-$ phenotype, this assay reports

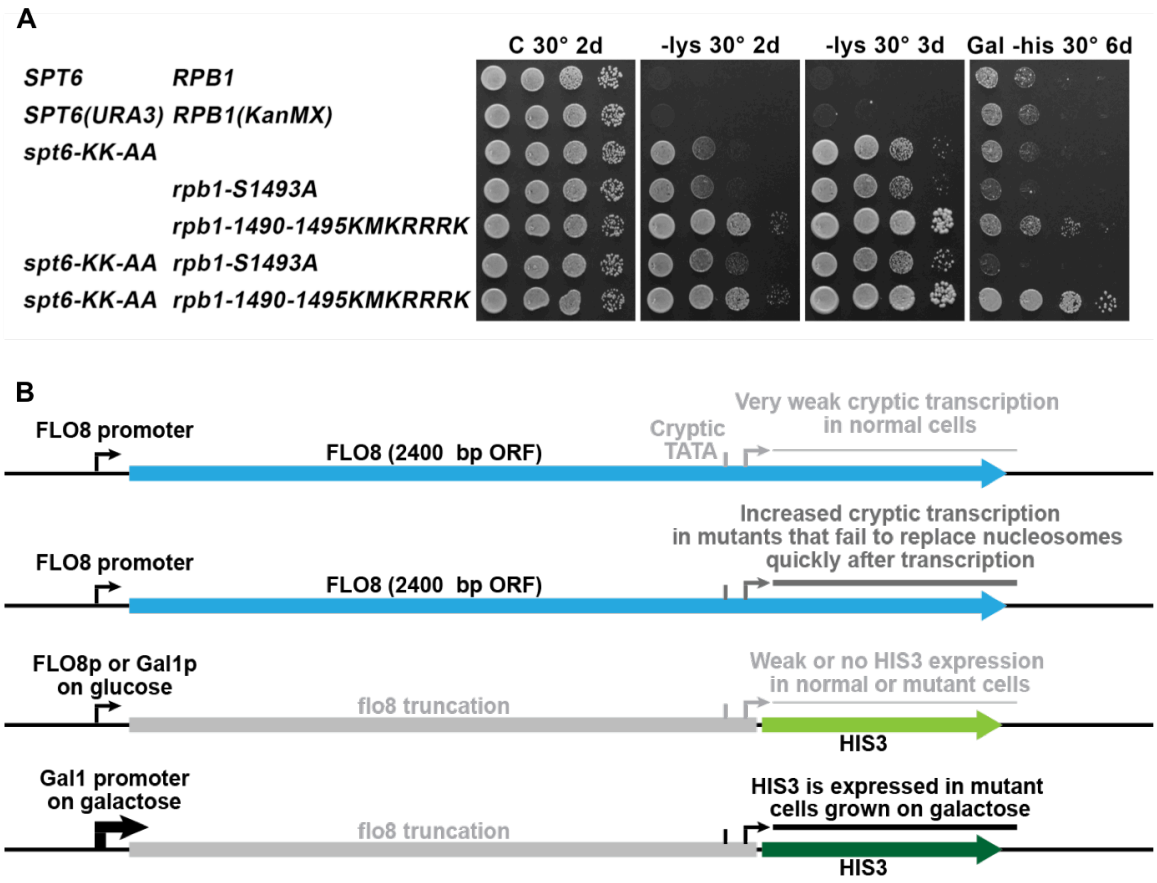



**Figure 3.7. The Spt6-Rpb1 interaction is important for maintaining repressive chromatin *in vivo*.**
A) Phenotypes of Spt6 and Rpb1 mutations integrated into the genomes of yeast at their native loci. *spt6-KK-AA* has alanine substitutions at K1355 and K1435 in the phosphoserine binding pocket. *rpb1-S1493A* has S1493 mutated to an alanine to prevent phosphorylation. *rpb1-1490-1495KMKRRRK* mutates residues $L^{1490}$MFSPLV$^{1496}$ to $K^{1490}$MKRRK$^{1496}$. Growth on medium lacking lysine (-lys) indicates the $Spt^-$ phenotype. Growth on medium containing galactose (Gal) and lacking histidine (-his) indicates activation of a cryptic promoter reporter.
B) Schematic of the cryptic promoter reporter. Normal FLO8 has a cryptic TATA site in the 3' end of the gene that is normally blocked by chromatin so it leads to very little transcription. Mutants that cause a defect in restoring chromatin to a normal repressive state after transcription of FLO8 lead to production of a short transcript containing only the 3' end of the gene. The HIS3 gene was inserted into *FLO8* downstream of the cryptic promoter and out of frame with the FLO8 ORF. In normal cells, this leads to little expression of HIS3, but some mutants with a very strong defect become His+ due to cryptic transcription. Enhanced transcription of the region by induction of the strong Gal1 promoter increases the probability of cryptic promoter activation, and leads to a His+ phenotype with less severe mutations. Growth on Gal -his therefore reveals a defect in restoration of nucleosomes after transcription by RNA Pol II. Figure kindly provided by Tim Formosa.

on chromatin integrity, but is specific to nucleosome reassembly during transcription (Figure 3.7B) (Cheung et al., 2008). In this yeast strain, the HIS3 gene is inserted downstream of a cryptic promoter in the *FLO8* gene. The cryptic promoter is normally repressed by chromatin, so these yeast should be unable to grow on media lacking histidine. Therefore, growth on media lacking histidine only when high levels of transcription are induced by the strong *GAL1* promoter driving the reporter reveals a defect in restoring nucleosomes following transcription by RNAPII. This phenotype is enhanced by *spt6-KK-AA*, suggesting that *rpb1-R3* retains partial Spt6 binding activity or these Spt6 residues have additional functions that contribute to chromatin maintenance during transcription. Collectively, these data validate the functional importance of the Spt6-Rpb1 interface that we have characterized biochemically and structurally, and implicate a role for the Spt6-Rpb1 interaction in reassembling repressive chromatin during transcription. Presumably, phosphorylation of Rpb1 S1493 recruits Spt6 to elongating RNAPII complexes where it can facilitate reassembly of repressive nucleosome structure following passage of the polymerase.

## Discussion

The Spt6 tSH2 domain comprises the only SH2 domains in yeast and is important for several aspects of Spt6 function including association with RNAPII. Spt6 tSH2 has been reported to mediate recruitment to RNAPII by directly binding the phosphorylated CTD. In contrast, we have shown that the CTD is neither necessary nor sufficient for Spt6 tSH2 binding. Rather, we find that Spt6 tSH2 binds the flexible linker that tethers the CTD to the core enzyme. Binding is

enhanced ~500 fold when Rpb1 S1493 is phosphorylated. This phosphorylation event is seen to occur *in vivo* because an antibody that specifically recognizes this modification detected the epitope in western blots performed on whole cell extracts. This residue and the two flanking residues are strictly conserved throughout evolution, further suggesting functional importance. Consistent with this implication, mutating the serine to alanine causes defects in transcriptional repression *in vivo*.

Our crystal structure revealed that the Rpb1 pS1493 fills a basic pocket formed by Spt6 residues K1355, Y1381, and K1435 of the non-canonical cSH2 domain. Mutation of K1435 to alanine reduced binding ~150 fold whereas mutations in the putative phosphotyrosine binding site had no effect. Yeast with mutations in the Spt6-Rpb1 interface are significantly impaired for transcriptional repression at promoters and in coding regions, suggesting that the interaction is important for nucleosome reassembly during transcription. Overall, these data suggest a model in which Rpb1 S1493 phosphorylation recruits Spt6 to elongation complexes where it is positioned to reassemble nucleosomes immediately following RNAPII passage.

To our knowledge, this is the first report of a specific function for the Rpb1 linker. The linker is required for viability in *S. cerevisiae* (Suh et al., 2013), and the only reported role for the linker is to stabilize association of the Rpb4/7 heterodimer with the core enzyme. This was inferred from a crystal structure of *S. pombe* RNAPII where the linker was substantially ordered due to contacts with Rpb4/7, but it is unclear if this interaction is functionally important. Interestingly,

our crystal structure reveals that the Spt6 tSH2 binding site on the linker overlaps with the region that contacts Rpb4/7. It is possible that the Spt6 tSH2-Rpb1 linker interaction contributes to transcriptional regulation by competing with the Rpb4/7-Rpb1 linker interaction, but currently there is no direct evidence to support this possibility.

Our identification of the authentic RNAPII-Spt6 binding site changes the interpretation of published data. In the previously reported model, Spt6 binds the phosphorylated CTD. However, in ChIP experiments, Spt6 occupancy across genes does not correlate with the abundance of individual CTD phosphorylation marks, but it does correlate with total RNAPII occupancy (Mayer et al., 2012; Mayer et al., 2010). This could be explained by the interaction with the Rpb1 linker if linker phosphorylation throughout the transcription cycle mimics Spt6 occupancy. Therefore, an important area of future study will be to characterize the prevalence of S1493 phosphorylation across transcription units. The antibody that we raised in this study will be an important tool for pursuing this goal.

While phosphorylation of the CTD has been well studied, this is the first report of a biological role for phosphorylation of the Rpb1 linker, and has important implications for how Spt6 recruitment to genes is regulated. Since phosphorylation has such a profound effect on binding, understanding the kinase(s) and phosphatase(s) that control this modification is essential. Likely candidates include the CTD kinases Ctk1 and Bur1 that contribute to Spt6 occupancy at the 5’ end of genes (Burugula et al., 2014), but others may exist.

The work presented in this chapter changes our understanding of how

Spt6 interacts with RNAPII. Identification of the authentic Sp6 tSH2 binding site on RNAPII has important implications for Spt6 function and the direction of future research. To our knowledge, this is the first binding interaction and biologically relevant phosphorylation event described for the Rpb1 linker. Collectively, these data provide strong support for a model in which Spt6 directly binds elongating RNAPII to facilitate co-transcriptional reassembly of nucleosomes following RNAPII passage.

## Materials and Methods

### Plasmids

Plasmid pCK859, expressing *S. cerevisiae* Rpb1 (YCp_LEU2_Rpop21), has been described previously (Suh et al., 2010). For plasmids expressing mutant Rpb1, PCR fragments containing the desired mutation were ligated into the Sph1 and Xho1 sites of pCK859. Mutant PCR fragments were produced using overlap extension PCR as described previously (Saiki et al., Nucleic Acid Research, 1988).

Plasmids expressing $His_6$ tagged Spt6 were described previously (Close et al., 2011). Plasmids expressing GST tagged Spt6 were produced by inserting the $Spt6^{1223\text{-}1451}$ coding sequence into the pDEST15 vector (Thermo Fisher Scientific). Mutations in the tSH2 domain were introduced using site-directed mutagenesis.

Yeast Strains

Yeast strain MAS015 was created by integrating a PreScission Protease Protein A tag into the C-terminus of the RPB3 gene in the protease-deficient yeast strain 7382 (from Tim Formosa). This strain was used for purification of wild type RNAPII.

Yeast strain 9138-4-2 (derived from CKY283, which has been described previously; Kaplan et al., 2008; Larson et al., 2012) contains a TAP tag at the C-terminus of the RPB3 gene, a disrupted genomic *RPB1*, and expresses Rpb1 from a plasmid (pRP112_YCp URA3). The strain was transformed with mutant Rbp1 plasmids, and loss of the *URA3* plasmid was selected by plating on 5-FOA (5-fluoroorotic acid). These strains were used to purify RNAPII with Rpb1 mutants.

Yeast strains used for assaying phenotypes were created and scored in the laboratory of Tim Formosa.

Protein Expression and Purification

*S. cerevisiae* RNAPII was purified from yeast strains containing either a PreScission Protease cleavable Protein A tag (wild type) or TAP tag (Rpb1 mutants) at the C-terminus of the Rpb3 subunit. RNAPII ΔCTD, Δlinker, and Rpb1 point mutants were purified from yeast expressing Rpb1 from a plasmid. Six liters of YPD (yeast extract peptone dextrose) were inoculated with 100 ml of a saturated overnight culture and incubated at 30°C for 2 days. Cells were harvested by centrifugation, washed once with cold water, and pelleted by centrifugation. Cells were frozen by passing through a syringe into liquid

nitrogen. Frozen cells were lysed under liquid nitrogen using a SPEX SamplePrep 6870 Freezer/Mill™.

Pulverized yeast were rehydrated and thawed in 1 pellet equivalent of lysis buffer (50 mM Tris pH 7.5, 500 mM KCl, 10% glycerol, 0.1% Tween-20, 10 μM $ZnCl_2$, 1 mM DTT, 2x pepstatin, 2x leupeptin, 2x aprotinin, 2x PMSF). Lysate was clarified by centrifugation at 37,000 RCF for 30 minutes. The supernatant was incubated with 2 ml of IgG sepharose resin (MP Biomedicals cappel™ antigen affinity gel) for 1 hour, followed by 6 washes with 4 ml of lysis buffer. Proteins were eluted by incubating with 4 ml of lysis buffer containing 100 μg PreScission Protease overnight at 4°C. In the morning, the supernatant was collected and the resin washed with 4 ml of lysis buffer. The supernatant and wash were pooled and diluted 4 fold with buffer QA (20 mM Tris-acetate pH 7.5, 0.5 mM EDTA, 10 μM $ZnCl_2$, 10% glycerol, 1 mM DTT). The sample was further purified via anion exchange chromatography (5 ml Hi-Trap Q column) using a gradient from 150 mM to 1,500 mM KOAc. The RNAPII peak fractions were collected, concentrated to 10 ml, and diluted 2 fold with buffer QA to reduce the salt concentration. RNAPII was exchanged into storage buffer (25 mM HEPES pH 7.5, 150 mM KOAc, 10 μM $ZnCl_2$, 10% glycerol, 5 mM DTT) by serial concentration and dilution. RNAPII was concentrated to ~5 μM and flash frozen in liquid nitrogen prior to storage at -80°C.

$Spt6^{1247\text{-}1451}$ proteins were expressed and purified as described previously (Close et al., 2011). GST-$Spt6^{1223\text{-}1451}$ expression and purification was similar with the exception that lysate was applied to glutathione agarose resin (Pierce)

instead of Ni-NTA resin.

Peptide Synthesis

Peptides were synthesized and purified by James Fulcher in the laboratory of Michael Kay.

Far Western Blots

Far western blots were performed as described previously (Wu et al., 2007). Briefly, 2.3 pmol of purified RNAPII was separated by SDS-PAGE on NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels (Invitrogen) electrophoresed in MOPS buffer, transferred to nitrocellulose membranes, and denatured in guanidine HCl. Proteins were renatured by stepwise incubation with protein binding buffer (20 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 10% glycerol, 0.1% Tween-20, 2% skim milk powder, 1 mM DTT) containing decreasing concentrations of guanidine HCl and incubated with blocking buffer [3% milk in TBS (25 mM Tris, 137 mM NaCl, 2.7 mM KCl)] for 1 hour at room temperature. Membranes were incubated with protein binding buffer containing 25 μM GST or GST-Spt6$^{1223\text{-}1451}$ overnight at 4°C. Membranes were washed 3 times with TBS-T (TBS with 0.05% Tween-20) and incubated with anti-GST antibody produced in goat (GE Healthcare 27-4577-01) diluted 1:1,000 in blocking buffer. After washing 3 times with TBS-T, membranes were incubated with a 1:10,000 dilution of infrared-labeled antisera that recognizes goat IgG. Bound proteins were detected using an Odyssey scanner.

Western Blots

Yeast whole cell extracts were prepared using the trichloroacetic acid (TCA) method as described previously (McCullough et al., 2015). Protein concentrations were determined using the Pierce™ BCA Protein Assay Kit, and 7 µM total protein was separated by SDS-PAGE. Rpb1 was detected with primary antibodies y-80 (Santa Cruz) or 8WG16 (Covance) that recognize the N-terminal region or CTD, respectively, and infrared-labeled secondary antibodies.

Polyclonal antisera that specifically recognizes Rpb1 pS1493 was produced by Covance. Briefly, rabbits were injected with the Rpb1 linker-derived peptide CDVKDELMF(pS)PLVDSGSN conjugated to KLH. The exsanguination bleed was subjected to positive and negative affinity purification steps over columns consisting of the phosphorylated peptide or unphosphorylated peptide respectively. The specificity for the phosphorylated peptide versus the unphosphorylated peptide was validated by EMSA.

Fluorescence Anisotropy

For direct binding assays, purified Spt6$^{1247\text{-}1451}$ was dialyzed overnight into anisotropy buffer (15 mM Tris pH 7.5, 100 mM NaCl, 5% glycerol, 0.5 mM EDTA pH 8.0, 2 mM BME) and titrated in 2 fold serial dilutions against a constant concentration (~0.5 nM) of fluorescently labeled peptide. Reactions were incubated for 20 minutes at 25°C and fluorescence anisotropy was measured using a BioTek Synergy Neo Microplate Reader set to 485 nm/535 nm excitation/emission wavelengths. $K_D$ values were determined by fitting the data in GraphPad Prim® using nonlinear least squared regression against $A = A_T / (2 \times$

[pep]) × (([pro] + [pep] + $K_D$) – (([pro] + [pep] + $K_D$)$^2$ – 4 × [pro] × [pep])$^{1/2}$) (LiCata and Wowor, 2008), where A is the measured anisotropy, $A_T$ is the total change in anisotropy, [pep] is the peptide concentration, and [pro] is the protein concentration.

For competition experiments, Spt6$^{1247\text{-}1451}$ and unlabeled peptides were dialyzed overnight into anisotropy buffer. The unlabeled peptide was titrated in 1.8 fold serial dilutions against constant concentrations of fluorescently labeled peptide (~0.5 nM) and Spt6$^{1247\text{-}1451}$ (200 nM). Reactions were incubated for 20 minutes at 25°C prior to measuring anisotropy. Kd values for the unlabeled ligand were determined by fitting the data in GraphPad Prims® using nonlinear least squared regression against the exact expression for competitive binding (Wang, 1995). The $K_D$ of the labeled ligand, the concentration of labeled ligand, and the concentration of protein were fixed.

X-ray Crystallography

86 μM Spt6$^{1247\text{-}1451}$ was mixed with a 2 fold molar excess of Rpb1$^{1476\text{-}1500\ pS1493}$ and concentrated. Crystals were grown at 4°C by vapor diffusion. Drops comprised 0.4 μl of 14 mg/ml protein and 0.2 μl well solution (40% ethanol, 100 mM Tris pH 7.0) (#7 of the Cryo II screen; Emerald BioSystems). Crystals were cooled by plunging into liquid nitrogen. Data were processed and scaled with HKL2000. The structure was determined by molecular replacement with Phaser (CCP4 program suite) using the coordinates of the yeast Spt6 tSH2 domain (PDB: 3PSJ) as the search model. Model building and refinement were performed in COOT (Emsley et al., 2010) and Phenix (Adams et al., 2010),

respectively. Figures of molecular structures were prepared using UCSF Chimera (Pettersen et al., 2004).

## References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W*., et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Adkins, M.W., and Tyler, J.K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. Mol Cell *21*, 405-416.

Andrulis, E.D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., and Lis, J.T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. Nature *420*, 837-841.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J *28*, 1067-1077.

Begum, N.A., Stanlie, A., Nakata, M., Akiyama, H., and Honjo, T. (2012). The histone chaperone Spt6 is required for activation-induced cytidine deaminase target determination through H3K4me3 regulation. J Biol Chem *287*, 32415-32429.

Bortvin, A., and Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. Science *272*, 1473-1476.

Bucheli, M.E., and Buratowski, S. (2005). Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. EMBO J *24*, 2150-2160.

Burckin, T., Nagel, R., Mandel-Gutfreund, Y., Shiue, L., Clark, T.A., Chong, J.L., Chang, T.H., Squazzo, S., Hartzog, G., and Ares, M., Jr. (2005). Exploring functional relationships between components of the gene expression machinery. Nat Struct Mol Biol *12*, 175-182.

Burugula, B.B., Jeronimo, C., Pathak, R., Jones, J.W., Robert, F., and Govind, C.K. (2014). Histone deacetylases and phosphorylated polymerase II C-terminal domain recruit Spt6 for cotranscriptional histone reassembly. Mol Cell Biol *34*, 4115-4129.

Cheung, V., Chua, G., Batada, N.N., Landry, C.R., Michnick, S.W., Hughes, T.R., and Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces

cerevisiae genome. PLoS Biol *6*, e277.

Clark-Adams, C.D., and Winston, F. (1987). The SPT6 gene is essential for growth and is required for delta-mediated transcription in Saccharomyces cerevisiae. Mol Cell Biol *7*, 679-686.

Close, D., Johnson, S.J., Sdano, M.A., McDonald, S.M., Robinson, H., Formosa, T., and Hill, C.P. (2011). Crystal structures of the S. cerevisiae Spt6 core and C-terminal tandem SH2 domain. J Mol Biol *408*, 697-713.

Compagnone-Post, P.A., and Osley, M.A. (1996). Mutations in the SPT4, SPT5, and SPT6 genes alter transcription of a subset of histone genes in Saccharomyces cerevisiae. Genetics *143*, 1543-1554.

DeGennaro, C.M., Alver, B.H., Marguerat, S., Stepanova, E., Davis, C.P., Bahler, J., Park, P.J., and Winston, F. (2013). Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. Mol Cell Biol *33*, 4779-4792.

Diebold, M.L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J., and Romier, C. (2010a). The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIS, Elongin A and Med26. EMBO J *29*, 3979-3991.

Diebold, M.L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., and Romier, C. (2010b). Noncanonical tandem SH2 enables interaction of elongation factor Spt6 with RNA polymerase II. J Biol Chem *285*, 38389-38398.

Dronamraju, R., and Strahl, B.D. (2014). A feed forward circuit comprising Spt6, Ctk1 and PAF regulates Pol II CTD phosphorylation and transcription elongation. Nucleic Acids Res *42*, 870-881.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev *113*, 8456-8490.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A.*, et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol Cell Biol *24*, 3324-3336.

Estruch, F., Peiro-Chova, L., Gomez-Navarro, N., Durban, J., Hodge, C., Del Olmo, M., and Cole, C.N. (2009). A genetic screen in Saccharomyces cerevisiae identifies new genes that interact with mex67-5, a temperature-sensitive allele of the gene encoding the mRNA export receptor. Mol Genet Genomics *281*, 125-134.

Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M., and Martens, J.A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. Genes Dev *25*, 29-40.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev *12*, 357-369.

Ivanovska, I., Jacques, P.E., Rando, O.J., Robert, F., and Winston, F. (2011). Control of chromatin structure by spt6: different consequences in coding and regulatory regions. Mol Cell Biol *31*, 531-541.

Johnson, S.J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S.L., and Hill, C.P. (2008). Crystal structure and RNA binding of the Tex protein from Pseudomonas aeruginosa. J Mol Biol *377*, 1460-1473.

Kaplan, C.D., Holland, M.J., and Winston, F. (2005). Interaction between transcription elongation factors and mRNA 3'-end formation at the Saccharomyces cerevisiae GAL10-GAL7 locus. J Biol Chem *280*, 913-922.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.

Kaplan, C.D., Larsson, K.M., and Kornberg, R.D. (2008). The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. Mol Cell *30*, 547-556.

Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. EMBO J *23*, 354-364.

Larson, M.H., Zhou, J., Kaplan, C.D., Palangat, M., Kornberg, R.D., Landick, R., and Block, S.M. (2012). Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. Proc Natl Acad Sci U S A *109*, 6555-6560.

Li, Y., and Kornberg, R.D. (1994). Interplay of positive and negative effectors in function of the C-terminal repeat domain of RNA polymerase II. Proc Natl Acad Sci U S A *91*, 2362-2366.

LiCata, V.J., and Wowor, A.J. (2008). Applications of fluorescence anisotropy to the study of protein-DNA interactions. Methods Cell Biol *84*, 243-262.

Liu, J., Zhang, J., Gong, Q., Xiong, P., Huang, H., Wu, B., Lu, G., Wu, J., and Shi, Y. (2011). Solution structure of tandem SH2 domains from Spt6 protein and their binding to the phosphorylated RNA polymerase II C-terminal domain. J Biol Chem *286*, 29218-29226.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science *336*, 1723-1725.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol *17*, 1272-1278.

McCullough, L., Connell, Z., Petersen, C., and Formosa, T. (2015). The abundant histone chaperones Spt6 and FACT collaborate to assemble, inspect, and maintain chromatin structure in Saccharomyces cerevisiae. Genetics *201*, 1031-1045.

McDonald, S.M., Close, D., Xin, H., Formosa, T., and Hill, C.P. (2010). Structure and biological importance of the Spn1-Spt6 interaction, and its regulatory role in nucleosome binding. Mol Cell *40*, 725-735.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217.

Okazaki, I.M., Okawa, K., Kobayashi, M., Yoshikawa, K., Kawamoto, S., Nagaoka, H., Shinkura, R., Kitawaki, Y., Taniguchi, H., Natsume, T.*, et al.* (2011). Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. Proc Natl Acad Sci U S A *108*, 7920-7925.

Perales, R., Erickson, B., Zhang, L., Kim, H., Valiquett, E., and Bentley, D. (2013). Gene promoters dictate histone occupancy within genes. EMBO J *32*, 2645-2656.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res *42*, W320-324.

Suh, H., Hazelbaker, D.Z., Soares, L.M., and Buratowski, S. (2013). The C-terminal domain of Rpb1 functions on other RNA polymerase II subunits. Mol Cell *51*, 850-858.

Suh, M.H., Meyer, P.A., Gu, M., Ye, P., Zhang, M., Kaplan, C.D., Lima, C.D., and Fu, J. (2010). A dual interface determines the recognition of RNA polymerase II by RNA capping enzyme. J Biol Chem *285*, 34027-34038.

Sun, M., Lariviere, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). J Biol Chem *285*, 41597-

41603.

Thebault, P., Boutin, G., Bhat, W., Rufiange, A., Martens, J., and Nourani, A. (2011). Transcription regulation by the noncoding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. Mol Cell Biol *31*, 1288-1300.

Wang, Z.X. (1995). An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. FEBS Lett *360*, 111-114.

Wu, Y., Li, Q., and Chen, X.Z. (2007). Detecting protein-protein interactions by Far western blotting. Nat Protoc *2*, 3278-3284.

Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes Dev *21*, 160-174.

Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev *22*, 3422-3434.

# CHAPTER 4

# IDENTIFICATION OF TOM1 AS A NOVEL BINDING PARTNER OF THE SPT6 TANDEM SH2 DOMAIN


## Summary

Spt6 is involved in several facets of gene expression including transcription elongation, chaperoning histones, and exporting mRNA. The C-terminus of Spt6 consists of a tSH2 domain comprised of the only SH2 domains predicted in yeast. The tSH2 domain contributes to Spt6 function and is necessary for maximal recruitment to sites of transcription. Binding interactions are likely central to Spt6 tSH2 function. The only published binding partner for Spt6 tSH2 is the phosphorylated CTD of RNAPII, but rigorous biochemistry suggests that this is a nonspecific interaction. In this work, we used pull-downs, mass spectrometry, and far western blotting to identify direct Spt6 tSH2 binding partners. We found that the tSH2 domain specifically binds the E3 ubiquitin ligase Tom1. This interaction requires Tom1 phosphorylation and was mapped to a central 30 residue fragment in a highly acidic region that competes with RNAPII for binding. Deletion of the Tom1-Spt6 binding region in *S. cerevisiae* suppresses temperature sensitivity caused by Tom1 overexpression, indicating that the interaction is biologically important. A cryo-EM reconstruction of Tom1 reveals a highly structured protein composed primarily of helical repeats. Overall, the data

presented in this chapter establish Tom1 as a biologically relevant binding partner of Spt6 tSH2 and lay a foundation for testing the functional importance of the interaction.

## Introduction

Spt6 is a highly conserved factor that is essential in yeast (Clark-Adams and Winston, 1987). Since its initial discovery in a genetic screen for factors required for normal transcription initiation (Winston et al., 1984), Spt6 has been implicated in several facets of gene expression including transcription elongation and termination, mRNA processing and export, posttranslational modification of histones, and nucleosome positioning (Ardehali et al., 2009; Bucheli and Buratowski, 2005; Clark-Adams and Winston, 1987; Compagnone-Post and Osley, 1996; DeGennaro et al., 2013; Dronamraju and Strahl, 2014; Endoh et al., 2004; Hartzog et al., 1998; Ivanovska et al., 2011; Kaplan et al., 2005; Perales et al., 2013; Yoh et al., 2007; Yoh et al., 2008). The diverse roles of Spt6 are important for a variety of biological processes such as embryogenesis in Zebrafish (Keegan et al., 2002; Kok et al., 2007), multiple stages of development in *Drosophila* (Ardehali et al., 2009), gut morphogenesis in *C. elegans* (Nishiwaki et al., 1993), and HIV transcriptional regulation (Vanti et al., 2009; Yoh et al., 2007).

An important role of Spt6 is to chaperone histones to reassemble nucleosomes in the wake of elongating RNAPII in order to repress aberrant transcription initiation (Adkins and Tyler, 2006; DeGennaro et al., 2013; Hainer et al., 2011; Ivanovska et al., 2011; Kaplan et al., 2003; Thebault et al., 2011). In

addition to its function as a histone chaperone, Spt6 is a transcription elongation factor that is capable of stimulating RNAPII elongation rates (Ardehali et al., 2009). This activity appears to be independent of Spt6 histone chaperone activity because Spt6 enhances RNAPII elongation rates *in vitro* on nucleosome-free DNA templates (Endoh et al., 2004; Yoh et al., 2007). Furthermore, Spt6 appears to coordinate the transcription cycle with mRNA processing and export because genetic and physical interactions have been reported between Spt6 and the mRNA processing and export machinery (Andrulis et al., 2002; Burckin et al., 2005; Estruch et al., 2009). Spt6 has several known activities that contribute to these diverse functions including binding to RNAPII, histones, DNA, nucleosomes, and Spn1 (Bortvin and Winston, 1996; Close et al., 2011; Diebold et al., 2010a; McCullough et al., 2015; McDonald et al., 2010; Yoh et al., 2007).

Spt6 is approximately 168 kDa and can be divided into three structural regions: an ~300 residue N-terminal region that is highly acidic and natively unstructured, a core that resembles the prokaryotic transcription factor *Tex*, and a C-terminal tSH2 domain (Close et al., 2011; Johnson et al., 2008). The tSH2 domain contains the only known SH2 domains in yeast and is conserved throughout eukaryotes (Close et al., 2011; Diebold et al., 2010b; Sun et al., 2010). Spt6 tSH2 is important for several facets of Spt6 function including maximal recruitment to genes, transcription elongation, maintenance of repressive chromatin, and mRNA processing and export (Burugula et al., 2014; Endoh et al., 2004; Hartzog et al., 1998; Mayer et al., 2010; Yoh et al., 2007). The only known activity of Spt6 tSH2 is direct binding to the phosphorylated

Rpb1 linker (discussed in Chapter 3), but this interaction had not been characterized when this work was initiated.

To better understand Spt6 tSH2 function, we took an unbiased approach to identify Spt6 tSH2 binding partners. We used mass spectrometry to identify factors that co-purify with Spt6 tSH2 over two affinity columns, and far western blots to identify direct binding partners. We identified the E3 ubiquitin ligase Tom1 as a novel Spt6 tSH2 binding partner. Tom1 is a 375 kDa protein that is homologous to mammalian Huwe1 (also called Huwe1/Mule/Lasu/Ureb1/Arf-BP1/HectH9/E3$^{Histone}$; Ptr1, Upl1, and Eel1) and has a C-terminal HECT domain. In humans, mutations in the gene encoding Huwe1 are associated with cancer and intellectual disability (Froyen et al., 2012; Froyen et al., 2008; Inoue et al., 2013; Isrie et al., 2013; Liu et al., 2012; Nava et al., 2012; Zhao et al., 2009). Tom1 has roles in mRNA processing and export, degradation of excess histones, transcription, and cell cycle regulation (Duncan et al., 2000; Iglesias et al., 2010; Kim and Koepp, 2012; Kim et al., 2012; Liu et al., 2005; Saleh et al., 1998; Singh et al., 2009; Utsugi et al., 1999). Our work shows that Spt6 directly binds a phosphorylated central 27 residue fragment of Tom1 that contributes to temperature sensitivity when Tom1 is overexpressed. Furthermore, we used cryo-EM to determine the structure of Tom1, and find that it folds into a single structure composed primarily of helical repeats.

## Results

### Spt6 Co-Purifies with RNAPII, Yra1, and Tom1

To identify factors that physically associate with Spt6 tSH2, we over-expressed tandem Strep-FLAG tagged Spt6 tSH2 (SF-Spt6$^{1223-1451}$) in yeast and performed tandem affinity purification (SF-TAP) (Gloeckner et al., 2007). The elution was resolved by SDS-PAGE and bands were excised for identification by mass spectrometry (Figure 4.1A). The background was very low, with no proteins apparent when the tandem affinity purification was performed from yeast transformed with an empty plasmid. The procedure was further validated by the strong co-purification of RNAPII, a known Spt6 tSH2 binding partner (Yoh et al., 2007). Many of the low molecular weight proteins were ribosomal proteins, which are common contaminants in this type of assay. In addition to ribosomal proteins, the 25 kDa mRNA export factor Yra1 and the 375 kDa E3 ubiquitin ligase Tom1 co-purified with SF-Spt6$^{1223-1451}$.

In order to determine if the associations with Yra1 and Tom1 were direct, we tested for Spt6 tSH2 binding in a far western blot. The SF-TAP elution samples were resolved by SDS-PAGE and transferred to a membrane. The membrane was probed with either purified recombinant GST-Spt6$^{1223-1451}$ or free GST and detected with antibodies against GST. As shown previously, Spt6$^{1223-1451}$ bound Rpb1, the largest subunit of RNAPII (Yoh et al., 2007). In addition, Spt6$^{1223-1451}$ also bound Tom1 (Figure 4.1B). Free GST does not bind, confirming that the interaction with Tom1 is specific to Spt6 tSH2 (Figure 4.1C). GST-Spt6$^{1223-1451}$ did not bind Yra1, suggesting that this interaction may be indirect

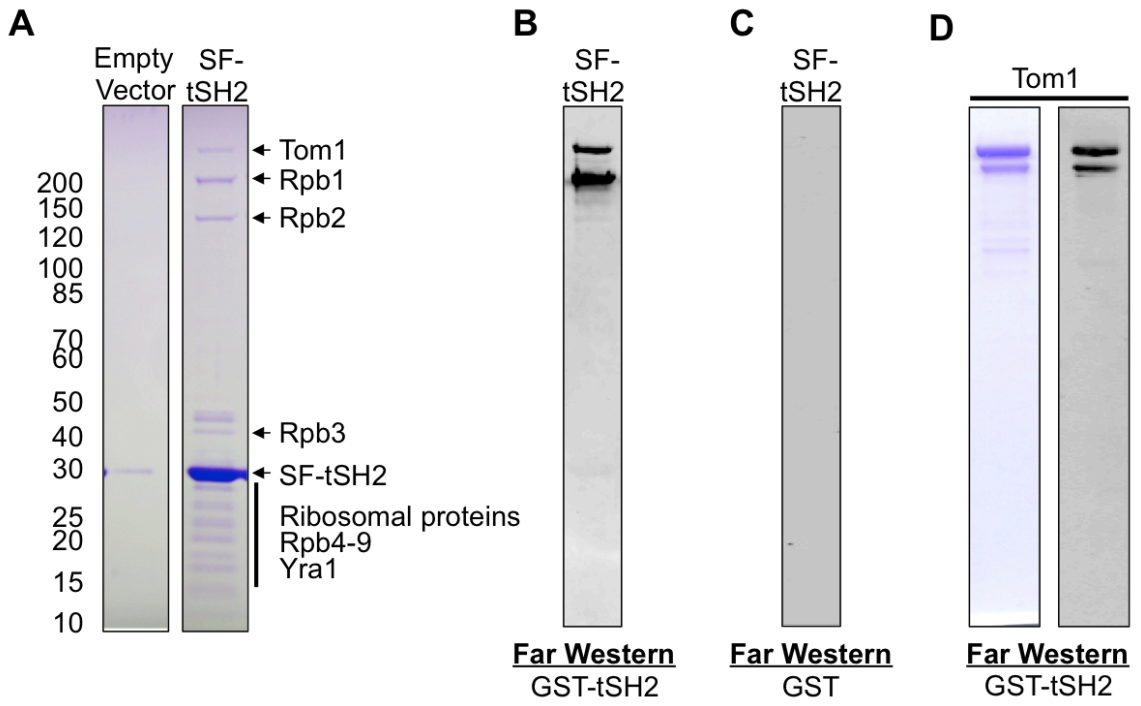



**Figure 4.1. Spt6 tSH2 binds Tom1**.
A) Coomassie stained gel of SF-TAP-purified SF-Spt6$^{1223\text{-}1451}$ expressed in yeast. Gel bands were excised and identified using mass spectrometry.
B) SF-TAP elutions were transferred to a membrane and used for far western blots with purified recombinant GST-Spt6$^{1223\text{-}1451}$ as a probe.
C) SF-TAP elutions were transferred to a membrane and used for far western blots with purified recombinant GST as a probe.
D) Tom1 was overexpressed in yeast and purified. Purified Tom1 was transferred to a membrane and used for far western blots with purified recombinant GST-Spt6$^{1223\text{-}1451}$ as a probe.

and is possibly mediated by mutual binding partners such as RNAPII or Tom1 (Iglesias et al., 2010; MacKellar and Greenleaf, 2011).

To verify the interaction with Tom1, we tested binding to highly purified Tom1. Tom1 with a C-terminal protein A tag was overexpressed in yeast under control of the GAL1-GAL10 promoter and was purified (Figure 4.1D). A faster migrating degradation product co-purified with full-length Tom1. Purified Tom1 was transferred to a membrane and tested for Spt6 tSH2 binding in a far western blot. Consistent with the far western blot using the SF-TAP sample, GST-Spt6$^{1223\text{-}1451}$ interacted with purified Tom1 (Figure 4.1D). Collectively, these results establish Tom1 as a previously unidentified Spt6 tSH2 binding partner.

Spt6 tSH2 Binds Phosphorylated Tom1

Since Spt6 tSH2 is known to bind phosphorylated proteins, we tested if the interaction with Tom1 is phosphorylation-dependent. Purified Tom1 was treated with calf intestinal alkaline phosphatase (CIP) prior to far western blotting with GST-Spt6$^{1223\text{-}1451}$. Treatment with CIP led to greatly reduced binding compared to untreated Tom1 (Figure 4.2B), indicating that phosphorylation contributes to binding. Because Spt6 tSH2 also binds phosphorylated Rpb1, we tested if Spt6 tSH2 binds Tom1 using the same binding pocket. The Spt6 K1435A mutation in the Spt6 cSH2 phosphoserine binding pocket disrupts the interaction with Rpb1, and this mutation eliminated binding to Tom1 in a far western blot (Figure 4.2B). These results suggest that Spt6 tSH2 binds phosphorylated Tom1 using the same binding interface as the Rpb1 linker.

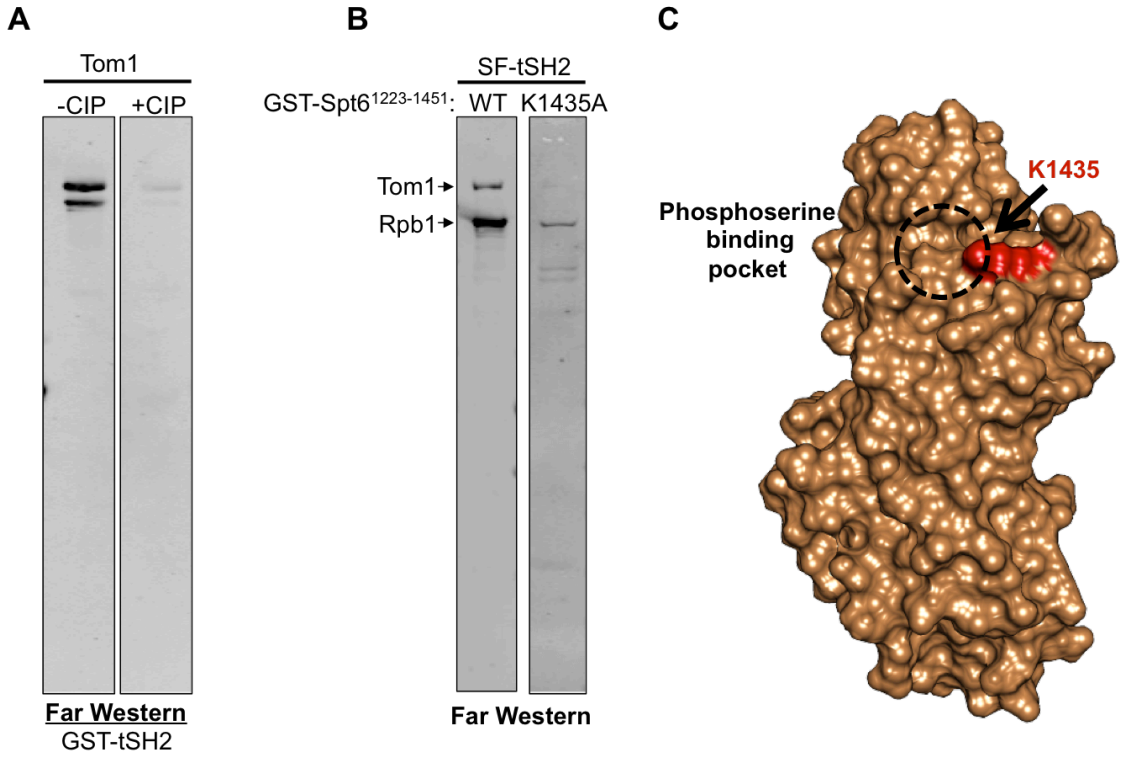



**Figure 4.2. Spt6 tSH2 binds phosphorylated Tom1 through the cSH2 domain phosphoserine binding pocket.**

A). Purified Tom1 was untreated (-CIP) or treated (+CIP) with CIP prior to SDS-PAGE and far western blotting with GST-Spt6$^{1223\text{-}1451}$. –CIP and +CIP were on the same gel and membrane, but intervening lanes were removed for clarity.

B) SF-TAP elutions were subjected to far western blotting with GST-Spt6$^{1223\text{-}1451}$ (WT) or GST-Spt6$^{1223\text{-}1451\ K1435A}$ (K1435A).

C) K1435A is a mutation in the cSH2 domain phosphoserine binding pocket that is important for binding Rpb1. Surface representation of the Spt6 tSH2 crystal structure (PDB ID: 3PSI) with the phosphoserine binding pocket highlighted with a dashed circle and K1435 colored red. Figure generated using UCSF Chimera (Pettersen et al., 2004).

Mapping the Tom1-Spt6 tSH2 Binding Site

Tom1 is a large protein (3,268 residues) with little predicted domain architecture (Figure 4.3A). To narrow down the region on Tom1 where Spt6 tSH2 binds, we subjected purified Tom1 to limited proteolysis and tested the resulting fragments for Spt6 tSH2 binding in a far western blot (Figure 4.3B). Despite having few recognizable domains, Tom1 is surprisingly resistant to proteolysis with an ~250 kDa fragment surviving up to 30 minutes in the presence of trypsin. This fragment retained binding to Spt6 tSH2 as did an ~60 kDa fragment whose appearance correlated with disappearance of the 250 kDa fragment. To map the 60 kDa fragment to the Tom1 sequence, we identified tryptic peptides using mass spectrometry. The majority of the peptides mapped to a central region of the Tom1 sequence that lacks recognizable structural motifs (Figure 4.3A). Because Tom1 binds the same Spt6 tSH2 interface as the Rpb1 linker, we reasoned that binding might be mediated through a similar sequence motif. By comparing the Spt6 tSH2-Rpb1 linker structure to Rpb1 sequence conservation, we identified three Rpb1 residues that contribute to the interface and are highly conserved (Figure 4.3C). Rpb1 L1490 is always a hydrophobic residue and F1492 and S1493 are strictly conserved. Given this insight, we searched Tom1 for the sequence Hb-x-F-S where Hb is any hydrophobic residue. Of the six candidate sequences in Tom1 that matched the motif, only one mapped to the sequence near the minimal binding fragment identified by mass spectrometry and was conserved throughout yeast (Figure 4.3D). Notably, the residue aligned with Rpb1 F1492 is invariant throughout the aligned sequences, and the residue

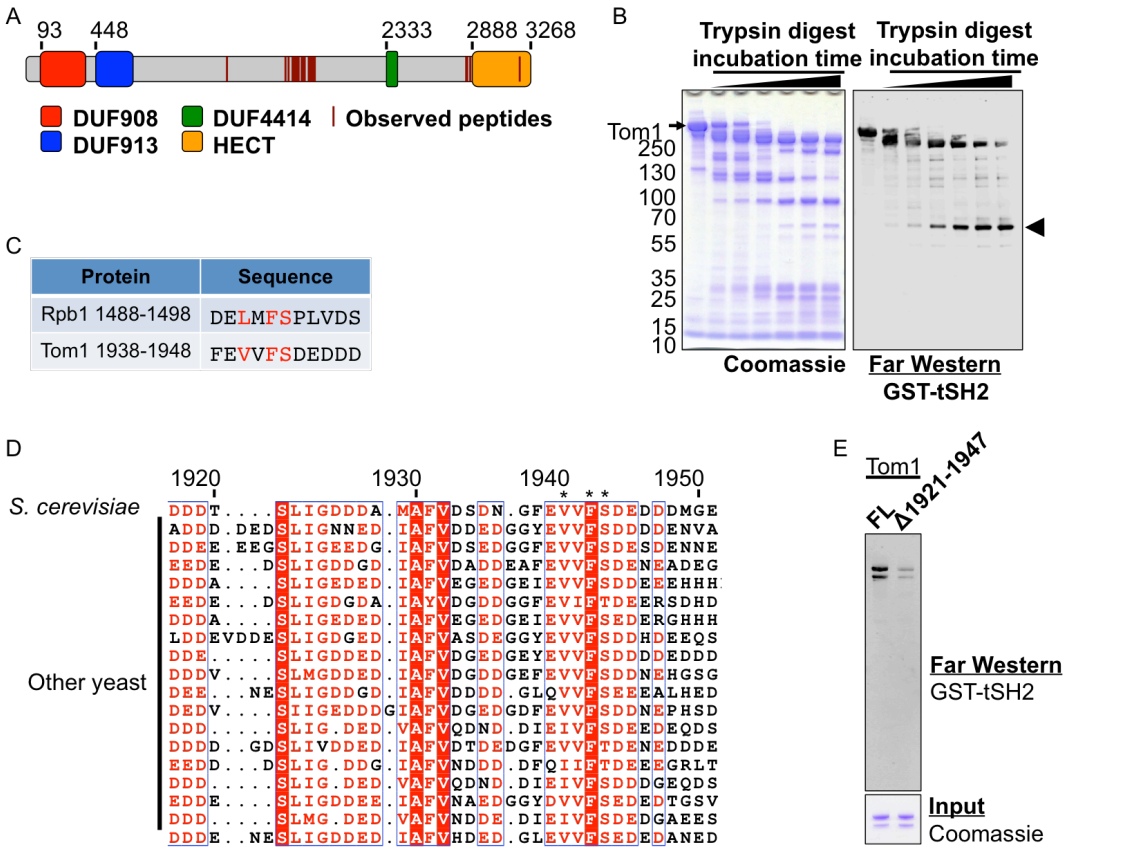



**Figure 4.3. Tom1 residues 1921-1947 are necessary for binding Spt6 tSH2.**
A) Schematic representation of Tom1 with predicted domains highlighted. Red lines represent peptides that were identified in the Spt6 binding fragment by mass spectrometry.
B) Purified Tom1 was partially proteolyzed by incubating with trypsin from 30 seconds to 30 minutes. Time points were resolved by SDS-PAGE and stained with coomassie or subjected to far western blotting with GST-Spt6$^{1223-1451}$ as a probe. Peptides in a ~60 kDa fragment that retained binding to Spt6 were identified using mass spectrometry (black arrowhead).
C) A candidate Tom1-Spt6 tSH2 binding site. The Tom1 sequence was searched for sequences that match the Rpb1 linker motif Hb-x-F-S, where Hb is any hydrophobic residue and x is any residue. Of the five hits, only one mapped to the Spt6 binding region, and was conserved throughout yeast
D) Alignment of the candidate Tom1-Spt6 tSH2 binding region (*S. cerevisiae* residues 1917-1951). Tom1 sequences from different yeast species were aligned using ConSurf (Ashkenazy et al., 2010) and ESPript (Robert and Gouet, 2014) was used for visualization. White characters on a red background represent invariant residues and red font on a white background represents highly similar (global similarity score $\geq$ 0.7) residues. Asterisks mark the residues matching the Hb-x-F-S motif highlighted in C.
E) Far western blot against full-length Tom1 (FL) and Tom1 Δ1921-1947 using GST-Spt6$^{1223-1451}$ as a probe.

aligned with Rpb1 S1493 is always a serine or a threonine, which can generally be phosphorylated like serine. To determine if this sequence is in fact the Spt6 tSH2 binding site, we tested binding to Tom1 lacking residues 1921-1947 (Tom1$^{\Delta 1921\text{-}1947}$), and found that GST-Spt6$^{12223\text{-}1451}$ binding was greatly reduced (Figure 4.3E). These results indicate that Tom1 residues 1921-1947 and phosphorylation, presumably of S1943, are necessary for binding Spt6 tSH2.

Tom1 1921-1951 Is Sufficient for Binding Spt6 tSH2

To determine if this region of Tom1 is sufficient for binding, we had a fluorescently labeled peptide spanning residues 1921-1951 synthesized with S1943 phosphorylated (Fl-Tom1$^{1921\text{-}1951\ pS1943}$). Fluorescence anisotropy was used to measure the binding affinity of purified recombinant Spt6$^{1247\text{-}1451}$ to Fl-Tom1$^{1921\text{-}1951\ pS1943}$. The binding constant ($K_D$) was 80 nM (Figure 4.4A), which is very similar to the $K_D$ observed for the Rpb1 linker peptide. Our assay was validated using a competition experiment in which unlabeled Tom1$^{1921\text{-}1951\ pS1943}$ peptide was used to compete with the labeled peptide for Spt6$^{1247\text{-}1451}$ binding. The inhibition constant ($K_I$) in this assay was 240 nM, which is in reasonably close agreement with the $K_D$ observed in direct binding assays (Figure 4.4B). This result indicates that the dye is not substantially interfering with or contributing to binding (Figure 4.4B). Because the Spt6 K1435A mutant reduced binding in the far western blot, we used fluorescence anisotropy to quantify the reduction in binding. Spt6$^{1247\text{-}1451\ K1435A}$ bound Tom1$^{1921\text{-}1951\ pS1943}$ with a $K_D$ of 5.6 μM, a reduction in affinity of 70 fold relative to wild type (Figure 4.4A).

Because the same Spt6 residues that are important for Tom1 binding are

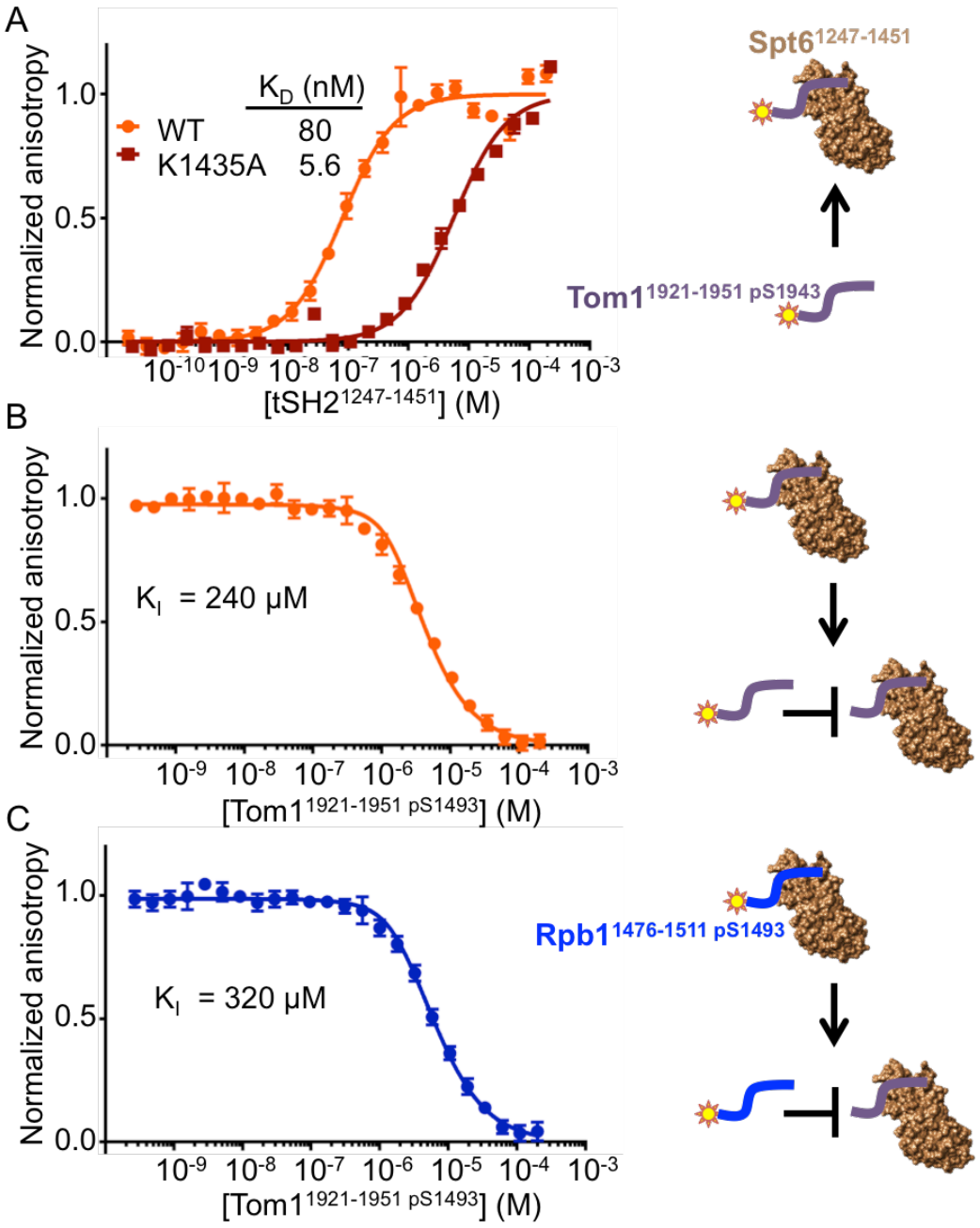



**Figure 4.4. Tom1 1921-1951 phosphorylated on S1493 is sufficient for binding Spt6 tSH2 and competes with the Rpb1 linker.**

A) Fluorescence anisotropy binding assay. Spt6$^{1247\text{-}1451}$ wild type (WT, orange circles) or K1435A (red squares) were titrated into Fl- Tom1$^{1921\text{-}1951\ pS1943}$.

B) Competition fluorescence anisotropy binding assay. Unlabeled Tom1$^{1921\text{-}1951\ pS1493}$ was titrated into constant concentrations of Spt6$^{1247\text{-}1451}$ (1 μM) and Fl-Tom1$^{1921\text{-}1951\ pS1943}$ (0.5 nM).

C) Competition fluorescence anisotropy binding assay. Unlabeled Tom1$^{1921\text{-}1951\ pS1493}$ was titrated into constant concentrations of Spt6$^{1247\text{-}1451}$ (1 μM) and Fl-Rpb1$^{1476\text{-}1511\ pS1493}$ (0.5 nM).

also important for RNAPII binding, we used a competition fluorescence anisotropy assay to determine if the interactions are competitive. Unlabeled Tom1$^{1921\text{-}1951\ pS1943}$ was titrated into constant concentrations of Spt6$^{1247\text{-}1451}$ and fluorescently labeled Rpb1$^{1476\text{-}1511\ pS1493A}$, and fluorescence anisotropy was measured. Tom1$^{1921\text{-}1951\ pS1943}$ competed with Rpb1$^{1476\text{-}1511\ pS1493A}$ for Spt6 tSH2 binding with a $K_I$ of 320 nM, consistent with the Ki determined using Fl-Tom1$^{1921\text{-}1951\ pS1943}$. These results indicate that Tom1 and RNAPII compete for Spt6 tSH2 binding and have similar affinities.

The Spt6 tSH2-Tom1 Interaction Is Important *In Vivo*

We used yeast genetics to determine if the Spt6 tSH2-Tom1 interaction is important *in vivo*. Tom1 deletion causes severe temperature sensitivity (Utsugi et al., 1999; Utsugi et al., 1995), but deletion of just the Tom1-Spt6 tSH2 binding domain (Tom1$^{\Delta 1926\text{-}1947}$) does not. Overexpressing Tom1 also causes a temperature sensitive phenotype, but overexpressing Tom1$^{\Delta 1926\text{-}1947}$, on the other hand, does not (data not shown); therefore, temperature sensitivity caused by Tom1 overexpression is likely related to Spt6 tSH2 binding, with one possibility being sequestration of Spt6 from RNAPII. Importantly, these results establish that the Spt6 interaction with Tom1 is relevant *in vivo*; however, at this point, the phenotypes do not hint at a specific functional role for the interaction.

Tom1 Structure

The resistance of Tom1 to proteolysis (Figure 4.4B) suggested that it is substantially folded. This was unexpected, as the four recognizable sequence

domains comprise less than 30% of the total sequence. The overall domain organization and large size of Tom1 has been conserved throughout evolution, suggesting that they are important features for Tom1 function. The specific domains of unknown function (DUF) seen in Tom1 are only found in Tom1 homologues, but no structural information is available. Notably, the sequences of DUF908 and DUF913 belong to a superfamily that contains helical repeat proteins. In order to better understand the structural and mechanistic details of Tom1 function, we set out to determine the structure.

Due to the large size of Tom1, we pursued single particle cryo-EM instead of x-ray crystallography. Tom1 purified from yeast eluted from a size exclusion column at an apparent molecular weight of ~480 kDa, which is consistent with a slightly elongated monomer (Figure 4.5A). Tom1 was imbedded in vitreous ice and imaged using cryo-EM, revealing numerous doughnut and horseshoe shaped particles (Figure 4.5B). 2D class averaging was used to discard bad particles and contaminants. The resulting good particles segregated into 2D class averages representative of multiple orientations with secondary structure features clearly visible (Figure 4.5C). 3D reconstruction with the clean data set yielded an 8 Å resolution structure of Tom1 that was sufficient to resolve alpha helices (Figure 4.5D and E).

The structure reveals an overall shape that resembles a lock washer consisting primarily of helical repeats. We were unable to dock the crystal structure of the HECT domain from Huwe1 (Pandya et al., 2010) into the density at either terminus, suggesting that it is not resolved in our reconstruction,

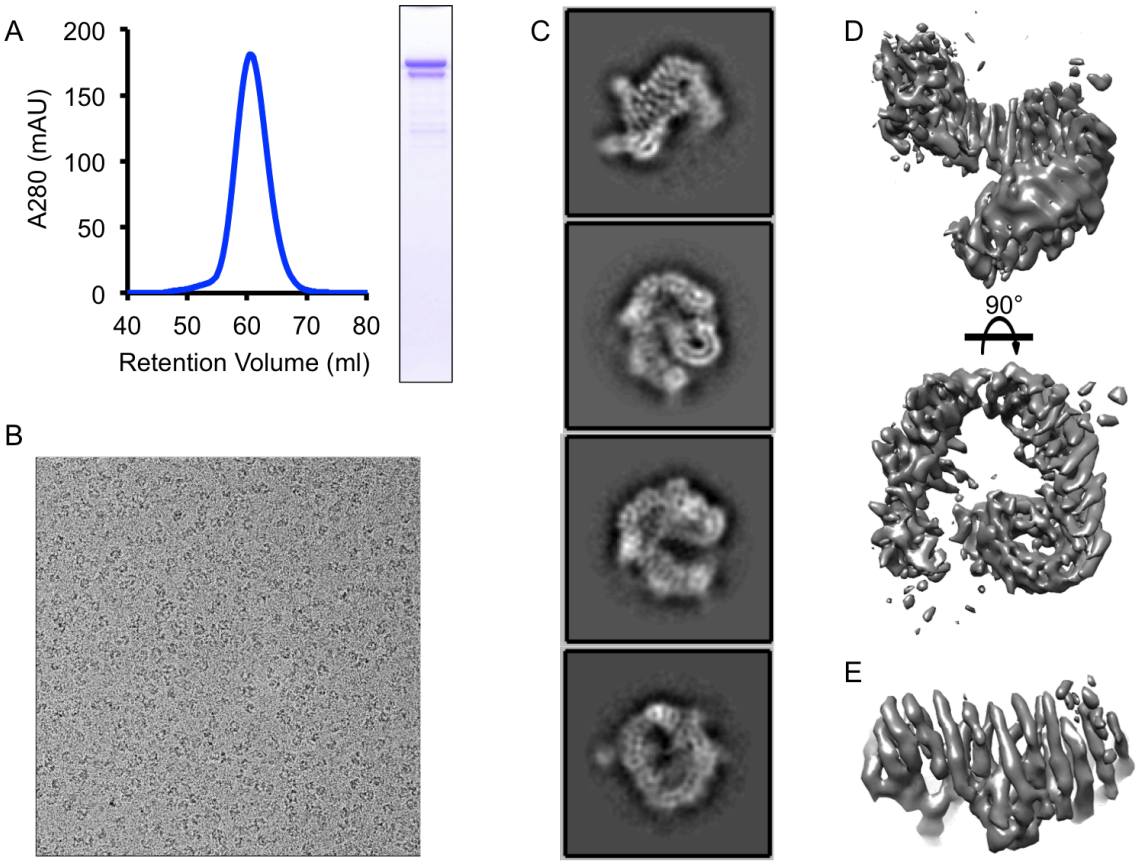



**Figure 4.5. 8 Å cryo-EM structure of Tom1.**
A) Size exclusion chromatogram of Tom1 injected on a Superdex 200 column. Tom1 elutes in a symmetric peak at a molecular weight of 480 kDa, consistent with an elongated monomer. Peak fractions were resolved using SDS-PAGE and stained with coomassie.
B) Purified Tom1 was frozen in vitreous ice and imaged using a TF20 microscope (FEI) and a K2 direct electron detector (Gatan).
C) Representative 2D class averages of Tom1.
D). Orthogonal views of an 8 Å Tom1 3D reconstruction.
E) A zoomed in view of the helical repeats showing resolution sufficient to see secondary structure features.
Figures D and E were generated using UCSF Chimera (Pettersen et al., 2004).

probably due to mobility relative to the rest of the structure. In addition, an ~300 residue region of Tom1 is highly acidic and predicted to be disordered, so it is unlikely to be resolved in the reconstruction. Excluding the HECT domain and the acidic region, there are ~305 kDa that could be structured. By comparing the volume of the Tom1 map to maps from published structures, we estimate that ~250-300 kDa of Tom1 are visible in our reconstruction, accounting for the majority of the Tom1 sequence except for the HECT domain and the acidic region. Therefore, despite minimal recognizable domains predicted by sequence, the majority of Tom1 is folded into a single, large structure.

## Discussion

We used SF-TAP followed by far western blotting and mass spectrometry to identify binding partners of the Spt6 tSH2 domain. Consistent with the work presented in Chapter 3, this analysis identified Rpb1 as an Spt6 tSH2 binding partner. Surprisingly, we found that the E3 ubiquitin ligase Tom1 also interacts directly with Spt6 tSH2. We further determined that Tom1 residues 1921-1947 are necessary for this interaction as is phosphorylation of S1943. Moreover, a phosphorylated Tom1 peptide binds Spt6 tSH2 with an affinity similar to Rpb1, and these interactions are competitive.

The biological relevance of the Spt6-Tom1 interaction is unclear, but our genetic data indicate that it is important. An attractive model is that the competition between RNAPII and Tom1 for Spt6 binding serves as a switch that signals the end of the transcription cycle. For example, the switch from RNAPII to Tom1 binding could be a checkpoint in mRNA maturation and export. Tom1 is

thought to ubiquitylate Yra1 in order to facilitate its removal from mRNA prior to nuclear export (Iglesias et al., 2010). Spt6 may recruit Tom1 to the nascent mRNA after transcription is complete, signaling Tom1 to ubiquitylate Yra1. This would ensure that Yra1 is only ubiquitylated when associated with a completely transcribed message, preventing nuclear export of truncated mRNAs. Consistent with this, Spt6 and Tom1 are each important for efficient nuclear export of mRNA and both associate with the mRNA export factor Yra1 (Estruch et al., 2009; Iglesias et al., 2010; Utsugi et al., 1999; Utsugi et al., 1995; Yoh et al., 2007). Testing this model will require monitoring changes in Yra1 ubiquitylation and defects in mRNA export upon deletion of the Spt6 binding region on Tom1.

Spt6 may also function with Tom1 in other processes such as transcriptional regulation, degradation of excess histones, and cell cycle regulation. Potential substrates for Tom1 that may be important for these processes include Spt7, histones, and Dia2, respectively. Monitoring ubiquitylation and turnover of these proteins when the Spt6-Tom1 interaction is impaired may provide insight into the functional role of binding. If a targeted approach does not provide clues pointing to a biological process, unbiased strategies may be necessary. These could include RNA-Seq to identify targets of Tom1 in transcriptional regulation, pull-downs and proteomics to identify binding partners, and ubiquitin-activated interaction traps (O'Connor et al., 2015) to identify potential Tom1 substrates.

In order to gain structural insight into the mechanism of Tom1, we determined its structure using single particle cryo-EM. Consistent with the

observation that Tom1 is resistant to proteolysis, the structure reveals that the majority of Tom1, excluding the HECT domain and an acidic region, are substantially folded. The bulk of the structure is comprised of helical repeats, which was somewhat expected because two of the DUFs belong to a superfamily of helical repeat proteins. At this point, it is unclear what the purpose of this unusual fold is, but conservation of the overall domain architecture and large size suggests a functional importance. The most attractive model is that the lock washer region is a binding platform that recruits Tom1 to different substrates. The focus of future work should be to determine a higher resolution structure of Tom1 and structures of co-complexes with binding partners including Spt6. These structures will be particularly insightful in light of connections between Huwe1 and human disease. The work presented here provides the foundation for pursuing additional structures.

These results reveal an unexpected phosphorylation dependent interaction between Spt6 and Tom1. This interaction may be the missing link that connects Spt6 to mRNA export, or it could be involved in other processes such as transcriptional regulation, histone degradation, and cell cycle regulation. Alternatively, Spt6 binding to Tom1 could be important for an unanticipated aspect of biology. Overall, this work provides a foundation to address these possibilities and to dissect the biological importance of the Spt6-Tom1 interaction.

## Materials and Methods

### Plasmids and Yeast Strains

Plasmid pMS1 expresses Spt6$^{1223-1451}$ with an N-terminal SF-TAP tag (Gloeckner et al., 2007) and a C-terminal SV40 nuclear localization signal under control of the GAL1 promoter. The coding sequence for Spt6$^{1223-1451}$ was PCR amplified from genomic DNA with a forward primer that introduced the sequence coding for the FLAG epitope. The PCR product was digested with NdeI and BamHI and ligated into pTF198 (Yep_LEU2_Gal1; Tim Formosa) to create pTF254. KpnI and XhoI sites were introduced between the FLAG epitope and the Spt6 tSH2 coding sequence using site-directed mutagenesis to create pTF254-KpnI-XhoI. Primers encoding a Twin-Strep-tag (IBA) flanked by KpnI and XhoI sites were annealed and ligated into digested pTF254-KpnI-XhoI to create pMS1. Other plasmids are described in Chapter 3.

Yeast strains 9750 and 9844 were constructed in the laboratory of Tim Formosa. Both strains express Tom1 with a C-terminal PreScission Protease site followed by a Protein A tag. Strain 9750 expresses full-length Tom1 whereas strain 9844 expresses Tom1 with residues 1921-1947 deleted. For both strains, the endogenous Tom1 promoter was replaced with the GAL1-GAL10 promoter in *S. cerevisiae*, allowing overexpression by addition of galactose to the media.

### Tandem Affinity Purification

Yeast transformed with plasmid pMS1 (SF-Spt6$^{1223-1451}$) or pTF198 (empty) were cultured in 2 liters synthetic media containing glycerol and lactate in baffled flasks at 30°C. When the OD$^{600}$ reached ~1, expression was induced by

the addition of galactose to 0.55% and the culture grown overnight. Cells were harvested by centrifugation and lysed under liquid nitrogen using a freezer mill. Subsequent steps were performed at 4°C. Yeast powder was suspended in an equal volume of lysis buffer (100 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 1 mM DTT, 0.1% Tween-20, 1x protease inhibitors [0.5 µg/ml aprotinin, 0.5 µg/ml leupeptin, 0.7 µg/ml pepstatin, 167 µg/ml PMSF], 1x phosphatase inhibitors [2 mM $NaVO_4$, 2 mM BGP, 2 mM NaF], 6 µg/ml DNAse, 2 µg/ml RNAse A) and clarified by centrifugation at 27,000 RCF for 30 minutes. 120 µg avidin was added to the clarified lysate and incubated for 15 minutes to block biotinylated proteins. Lysates were clarified further by ultracentrifugation at 107,000 RCF for 1 hour. The supernatant was applied to 100 µl Strep-Tactin® Superflow resin (IBA) and incubated for 1 hour. The resin was washed with 10 column volumes of wash buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol, 0.1% Tween-20, 1x protease and phosphatase inhibitors, 6 µg/ml DNAse, 2 µg/ml RNAse A) and eluted with buffer containing 2 mM desthiobiotin. The eluate was applied to 50 µl anti-FLAG® M2 affinity gel (Sigma-Aldrich) and incubated for 1 hour. The resin was washed 3 times with 100 µl wash buffer and eluted with 150 µl wash buffer containing 200 µg/ml FLAG peptide. Proteins were separated by SDS-PAGE on NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels (Invitrogen) in MOPS running buffer, and proteins were visualized by coomassie staining.

Protein Expression and Purification

Spt6 proteins were purified as described in Chapter 3.

Tom1 was purified from yeast strains 9750 and 9844. Cultures that were

grown to saturation were used to inoculate 2 L of YP containing glycerol and lactate to an $OD^{600}$ of 0.2. Yeast were grown at 30°C in baffled flasks on an orbital shaker. Expression was induced by the addition of galactose to ~0.5% when the $OD^{600}$ reached ~1, and cultures were grown overnight. Cells were harvested by centrifugation and lysed under liquid nitrogen using a freezer mill. Subsequent steps were performed at 4°C. Yeast powder was suspended in 2 pellet volumes of suspension buffer (50 mM Tris pH 7.5, 500 mM NaCl, 10% glycerol, 1 mM DTT, 0.1% Tween-20, 1x protease inhibitors) and clarified by centrifugation at 37,000 RCF for 30 minutes. Clarified lysate was incubated with 4 ml IgG Sepharose 6 Fast Flow resin (GE Healthcare) for 1 hour. Resin was washed 5 times with 8 ml of suspension buffer and 5 times with 4 ml IgG wash buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol, 1 mM DTT, 1x protease inhibitors). Tom1 was eluted by adding 4 ml IgG wash buffer containing 24 μg/ml PreScission Protease and incubating overnight. The resin was washed with an additional 8 ml IgG wash buffer that was pooled with the elution. The IgG elution was applied to a 5 ml HiTrap Q column (GE Healthcare) and eluted with a NaCl gradient. Tom1 eluted at ~500 mM NaCl. Fractions containing Tom1 were pooled, concentrated, and loaded onto a Superdex 200 size exclusion column (GE Healthcare) equilibrated in sizing buffer (15 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 2 mM BME).

Protein Identification

Krishna Parsawar at the University of Utah Metabolomics, Proteomics and Mass Spectrometry Core Facilities performed protein identification by mass

spectrometry.

Far Western Blots

Far western blots were performed as described previously (Wu et al., 2007) and in Chapter 3. Either 20 µl FLAG elution or 1.1 pmol purified Tom1 were loaded into each lane of a NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gel (Invitrogen) and run in MOPS buffer. For CIP treated Tom1, purified Tom1 was diluted to 300 nM in phosphatase buffer (50 mM Tris pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 5% glycerol, 0.1% Tween-20, 1 mM DTT, 1x protease inhibitors) and incubated with 0.3 U/µl CIP (New England BioLabs) for 1 hour at room temperature prior to loading onto a gel.

Limited Proteolysis

100 µg of purified Tom1 was incubated with 100 ng trypsin in 200 µl sizing buffer at room temperature. Time points were collected at 30 seconds, 1 minute, 3 minutes, 10 minutes, 20 minutes, and 30 minutes. Reactions were quenched by the addition of aprotinin, leupeptin, PMSF, and SDS-loading dye followed by boiling for 3 minutes. Time points were resolved by SDS-PAGE on NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels (Invitrogen) in MOPS running buffer, and proteins were visualized by coomassie staining or transferred to a membrane for far western blotting.

Fluorescence Anisotropy

Fluorescence anisotropy assays were performed as described in Chapter 3 Materials and Methods.

Cryo-EM

3 µl of purified Tom1 at 2.5 µM was applied to glow discharged Quantifoil R2/2 holey carbon grids and incubated for 30 seconds at 4°C and 80% humidity. Grids were blotted for 8 seconds and plunge frozen in liquid ethane using a Vitrobot (FEI). Data were collected on a TF20 (FEI) microscope operating at 200kV equipped with a K2 Summit direct electron detector (Gatan).

CTF parameters were estimated using CTFFIND4 (Rohou and Grigorieff, 2015). Initial 2D class averages were calculated from ~9,000 particles that were picked semi-automatically using the *swarm* tool in e2boxer.py (EMAN2) (Tang et al., 2007). Eight distinct 2D class averages were low pass filtered to 20 Å and used as templates for autopicking in RELION (Scheres, 2015). The resulting particles were filtered in multiple rounds of 2D class averaging. A consensus 3D reconstruction was computed and refined in RELION (Scheres, 2012) using a starting model generated from low pass filtered 2D class averages in the EMAN2 package.

## References

Adkins, M.W., and Tyler, J.K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. Mol Cell *21*, 405-416.

Andrulis, E.D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., and Lis, J.T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. Nature *420*, 837-841.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J *28*, 1067-1077.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf

2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res *38*, W529-533.

Bortvin, A., and Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. Science *272*, 1473-1476.

Bucheli, M.E., and Buratowski, S. (2005). Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. EMBO J *24*, 2150-2160.

Burckin, T., Nagel, R., Mandel-Gutfreund, Y., Shiue, L., Clark, T.A., Chong, J.L., Chang, T.H., Squazzo, S., Hartzog, G., and Ares, M., Jr. (2005). Exploring functional relationships between components of the gene expression machinery. Nat Struct Mol Biol *12*, 175-182.

Burugula, B.B., Jeronimo, C., Pathak, R., Jones, J.W., Robert, F., and Govind, C.K. (2014). Histone deacetylases and phosphorylated polymerase II C-terminal domain recruit Spt6 for cotranscriptional histone reassembly. Mol Cell Biol *34*, 4115-4129.

Clark-Adams, C.D., and Winston, F. (1987). The SPT6 gene is essential for growth and is required for delta-mediated transcription in Saccharomyces cerevisiae. Mol Cell Biol *7*, 679-686.

Close, D., Johnson, S.J., Sdano, M.A., McDonald, S.M., Robinson, H., Formosa, T., and Hill, C.P. (2011). Crystal structures of the S. cerevisiae Spt6 core and C-terminal tandem SH2 domain. J Mol Biol *408*, 697-713.

Compagnone-Post, P.A., and Osley, M.A. (1996). Mutations in the SPT4, SPT5, and SPT6 genes alter transcription of a subset of histone genes in Saccharomyces cerevisiae. Genetics *143*, 1543-1554.

DeGennaro, C.M., Alver, B.H., Marguerat, S., Stepanova, E., Davis, C.P., Bahler, J., Park, P.J., and Winston, F. (2013). Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. Mol Cell Biol *33*, 4779-4792.

Diebold, M.L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J., and Romier, C. (2010a). The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIS, Elongin A and Med26. EMBO J *29*, 3979-3991.

Diebold, M.L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., and Romier, C. (2010b). Noncanonical tandem SH2 enables interaction of elongation factor Spt6 with RNA polymerase II. J Biol Chem *285*, 38389-38398.

Dronamraju, R., and Strahl, B.D. (2014). A feed forward circuit comprising Spt6, Ctk1 and PAF regulates Pol II CTD phosphorylation and transcription elongation. Nucleic Acids Res *42*, 870-881.

Duncan, K., Umen, J.G., and Guthrie, C. (2000). A putative ubiquitin ligase required for efficient mRNA export differentially affects hnRNP transport. Curr Biol *10*, 687-696.

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A.*, et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol Cell Biol *24*, 3324-3336.

Estruch, F., Peiro-Chova, L., Gomez-Navarro, N., Durban, J., Hodge, C., Del Olmo, M., and Cole, C.N. (2009). A genetic screen in Saccharomyces cerevisiae identifies new genes that interact with mex67-5, a temperature-sensitive allele of the gene encoding the mRNA export receptor. Mol Genet Genomics *281*, 125-134.

Froyen, G., Belet, S., Martinez, F., Santos-Reboucas, C.B., Declercq, M., Verbeeck, J., Donckers, L., Berland, S., Mayo, S., Rosello, M.*, et al.* (2012). Copy-number gains of HUWE1 due to replication- and recombination-based rearrangements. Am J Hum Genet *91*, 252-264.

Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C., Bauters, M., Govaerts, K., Vandeleur, L., Van Esch, H.*, et al.* (2008). Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. Am J Hum Genet *82*, 432-443.

Gloeckner, C.J., Boldt, K., Schumacher, A., Roepman, R., and Ueffing, M. (2007). A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. Proteomics *7*, 4228-4234.

Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M., and Martens, J.A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. Genes Dev *25*, 29-40.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev *12*, 357-369.

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C., and Stutz, F. (2010). Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev *24*, 1927-1938.

Inoue, S., Hao, Z., Elia, A.J., Cescon, D., Zhou, L., Silvester, J., Snow, B., Harris, I.S., Sasaki, M., Li, W.Y.*, et al.* (2013). Mule/Huwe1/Arf-BP1 suppresses Ras-driven tumorigenesis by preventing c-Myc/Miz1-mediated down-regulation of p21 and p15. Genes Dev *27*, 1101-1114.

Isrie, M., Kalscheuer, V.M., Holvoet, M., Fieremans, N., Van Esch, H., and

Devriendt, K. (2013). HUWE1 mutation explains phenotypic severity in a case of familial idiopathic intellectual disability. Eur J Med Genet *56*, 379-382.

Ivanovska, I., Jacques, P.E., Rando, O.J., Robert, F., and Winston, F. (2011). Control of chromatin structure by spt6: different consequences in coding and regulatory regions. Mol Cell Biol *31*, 531-541.

Johnson, S.J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S.L., and Hill, C.P. (2008). Crystal structure and RNA binding of the Tex protein from Pseudomonas aeruginosa. J Mol Biol *377*, 1460-1473.

Kaplan, C.D., Holland, M.J., and Winston, F. (2005). Interaction between transcription elongation factors and mRNA 3'-end formation at the Saccharomyces cerevisiae GAL10-GAL7 locus. J Biol Chem *280*, 913-922.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.

Keegan, B.R., Feldman, J.L., Lee, D.H., Koos, D.S., Ho, R.K., Stainier, D.Y., and Yelon, D. (2002). The elongation factors Pandora/Spt6 and Foggy/Spt5 promote transcription in the zebrafish embryo. Development *129*, 1623-1632.

Kim, D.H., and Koepp, D.M. (2012). Hect E3 ubiquitin ligase Tom1 controls Dia2 degradation during the cell cycle. Mol Biol Cell *23*, 4203-4211.

Kim, D.H., Zhang, W., and Koepp, D.M. (2012). The Hect domain E3 ligase Tom1 and the F-box protein Dia2 control Cdc6 degradation in G1 phase. J Biol Chem *287*, 44212-44220.

Kok, F.O., Oster, E., Mentzer, L., Hsieh, J.C., Henry, C.A., and Sirotkin, H.I. (2007). The role of the SPT6 chromatin remodeling factor in zebrafish embryogenesis. Dev Biol *307*, 214-226.

Liu, Y.X., Zhang, S.F., Ji, Y.H., Guo, S.J., Wang, G.F., and Zhang, G.W. (2012). Whole-exome sequencing identifies mutated PCK2 and HUWE1 associated with carcinoma cell proliferation in a hepatocellular carcinoma patient. Oncol Lett *4*, 847-851.

Liu, Z., Oughtred, R., and Wing, S.S. (2005). Characterization of E3Histone, a novel testis ubiquitin protein ligase which ubiquitylates histones. Mol Cell Biol *25*, 2819-2831.

MacKellar, A.L., and Greenleaf, A.L. (2011). Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. J Biol Chem *286*, 36385-36395.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription

complex. Nat Struct Mol Biol *17*, 1272-1278.

McCullough, L., Connell, Z., Petersen, C., and Formosa, T. (2015). The abundant histone chaperones Spt6 and FACT collaborate to assemble, inspect, and maintain chromatin structure in Saccharomyces cerevisiae. Genetics *201*, 1031-1045.

McDonald, S.M., Close, D., Xin, H., Formosa, T., and Hill, C.P. (2010). Structure and biological importance of the Spn1-Spt6 interaction, and its regulatory role in nucleosome binding. Mol Cell *40*, 725-735.

Nava, C., Lamari, F., Heron, D., Mignot, C., Rastetter, A., Keren, B., Cohen, D., Faudet, A., Bouteiller, D., Gilleron, M.*, et al.* (2012). Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. Transl Psychiatry *2*, e179.

Nishiwaki, K., Sano, T., and Miwa, J. (1993). emb-5, a gene required for the correct timing of gut precursor cell division during gastrulation in Caenorhabditis elegans, encodes a protein similar to the yeast nuclear protein SPT6. Mol Gen Genet *239*, 313-322.

O'Connor, H.F., Lyon, N., Leung, J.W., Agarwal, P., Swaim, C.D., Miller, K.M., and Huibregtse, J.M. (2015). Ubiquitin-Activated Interaction Traps (UBAITs) identify E3 ligase binding partners. EMBO Rep *16*, 1699-1712.

Pandya, R.K., Partridge, J.R., Love, K.R., Schwartz, T.U., and Ploegh, H.L. (2010). A structural element within the HUWE1 HECT domain modulates self-ubiquitylation and substrate ubiquitylation activities. J Biol Chem *285*, 5664-5673.

Perales, R., Erickson, B., Zhang, L., Kim, H., Valiquett, E., and Bentley, D. (2013). Gene promoters dictate histone occupancy within genes. EMBO J *32*, 2645-2656.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res *42*, W320-324.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol *192*, 216-221.

Saleh, A., Collart, M., Martens, J.A., Genereaux, J., Allard, S., Cote, J., and Brandl, C.J. (1998). TOM1p, a yeast hect-domain protein which mediates transcriptional regulation through the ADA/SAGA coactivator complexes. J Mol Biol *282*, 933-946.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol *180*, 519-530.

Scheres, S.H. (2015). Semi-automated selection of cryo-EM particles in RELION-1.3. J Struct Biol *189*, 114-122.

Singh, R.K., Kabbaj, M.H., Paik, J., and Gunjan, A. (2009). Histone levels are regulated by phosphorylation and ubiquitylation-dependent proteolysis. Nat Cell Biol *11*, 925-933.

Sun, M., Lariviere, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). J Biol Chem *285*, 41597-41603.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol *157*, 38-46.

Thebault, P., Boutin, G., Bhat, W., Rufiange, A., Martens, J., and Nourani, A. (2011). Transcription regulation by the noncoding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. Mol Cell Biol *31*, 1288-1300.

Utsugi, T., Hirata, A., Sekiguchi, Y., Sasaki, T., Toh-e, A., and Kikuchi, Y. (1999). Yeast tom1 mutant exhibits pleiotropic defects in nuclear division, maintenance of nuclear structure and nucleocytoplasmic transport at high temperatures. Gene *234*, 285-295.

Utsugi, T., Toh-e, A., and Kikuchi, Y. (1995). A high dose of the STM1 gene suppresses the temperature sensitivity of the tom1 and htr1 mutants in Saccharomyces cerevisiae. Biochim Biophys Acta *1263*, 285-288.

Vanti, M., Gallastegui, E., Respaldiza, I., Rodriguez-Gil, A., Gomez-Herreros, F., Jimeno-Gonzalez, S., Jordan, A., and Chavez, S. (2009). Yeast genetic analysis reveals the involvement of chromatin reassembly factors in repressing HIV-1 basal transcription. PLoS Genet *5*, e1000339.

Winston, F., Chaleff, D.T., Valent, B., and Fink, G.R. (1984). Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics *107*, 179-197.

Wu, Y., Li, Q., and Chen, X.Z. (2007). Detecting protein-protein interactions by Far western blotting. Nat Protoc *2*, 3278-3284.

Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes Dev *21*, 160-174.

Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev *22*, 3422-3434.

Zhao, X., D, D.A., Lim, W.K., Brahmachary, M., Carro, M.S., Ludwig, T., Cardo, C.C., Guillemot, F., Aldape, K., Califano, A.*, et al.* (2009). The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. Dev Cell *17*, 210-221.

# CHAPTER 5

# MECHANISM OF SPT6 HISTONE CHAPERONE ACTIVITY AND ITS REGULATION BY SPN1


## Summary

Spt6 is a histone chaperone that binds and reassembles nucleosomes in the wake of elongating RNAPII. This activity is essential to restore the default repressive chromatin state and to prevent inappropriate transcription initiation. Nucleosome binding by Spt6 requires a region of the protein that overlaps with the Spn1-binding site, and Spn1 competes with nucleosomes for Spt6 binding *in vitro*. In an effort to understand the molecular details of the histone chaperone activity of Spt6, we have performed direct binding and competition assays. These data reveal that Spt6 binds the globular core of histone H3-H4 tetramers [$(H3\text{-}H4)_2$] and competes with DNA for binding. Further, we provide evidence that the Spn1-Spt6 and Spt6-$(H3\text{-}H4)_2$ interactions are mutually exclusive. Mutations that specifically disrupt the interaction with $(H3\text{-}H4)_2$ *in vitro* cause synthetic growth defects *in vivo* when combined with mutations in Spn1 or other histone chaperones. Overall, the data presented here support the model that Spn1 plays a regulatory role in the histone chaperone activity of Spt6, and imply a mechanism by which Spt6 chaperones histones by direct competition with DNA.

## Introduction

Eukaryotic organisms use an efficient packaging system known as chromatin to compact large genomes into the nucleus of each cell (Li and Reinberg, 2011). However, much of the DNA needs to be accessed on a regular basis, which requires that chromatin compaction to be dynamic. For example, transcription by RNA polymerases requires accessing, processing, and repackaging relevant segments of the genome. Therefore, chromatin dynamics are intimately linked to biological function and are highly regulated.

The fundamental packaging unit of chromatin is the nucleosome, which consists of two copies of each of the core histone proteins H2A, H2B, H3, and H4 that are wrapped 1.7 times by ~147 base pairs of DNA (Luger et al., 1997). Nucleosomes create a physical barrier to DNA templated processes such as transcription by RNAPII (Lorch et al., 1992). Consequently, different classes of proteins and protein complexes mediate access to DNA by altering nucleosome position and occupancy. Histone chaperones bind the highly charged surfaces of histones to prevent aberrant interactions and promote assembly or disassembly of nucleosomes (Gurard-Levin et al., 2014; Park and Luger, 2008; Ransom et al., 2010). Spt6 is a highly conserved, essential histone chaperone that is involved in several facets of gene expression including transcription, mRNA processing and export, histone posttranslational modification, and nucleosome positioning (Ardehali et al., 2009; Bucheli and Buratowski, 2005; Clark-Adams and Winston, 1987; Compagnone-Post and Osley, 1996; DeGennaro et al., 2013; Dronamraju and Strahl, 2014; Endoh et al., 2004; Hartzog et al., 1998; Ivanovska et al., 2011;

Kaplan et al., 2005; Perales et al., 2013; Yoh et al., 2007; Yoh et al., 2008).

The most well-characterized role of Spt6 is to reassemble nucleosomes in the wake of RNAPII in order to re-establish a repressive chromatin state that prevents initiation from cryptic promoters (Kaplan et al., 2003). Spt6 coordinates nucleosome reassembly with transcription by directly binding RNAPII (Yoh et al., 2007). Other Spt6 activities that contribute to nucleosome reassembly include direct binding to histones (Bortvin and Winston, 1996; McCullough et al., 2015), nucleosomes (McDonald et al., 2010), and DNA (Close et al., 2011).

The *Saccharomyces cerevisiae* Spt6 protein is large (~168 kDa) and has at least three distinct structural regions (Close et al., 2011). The disordered N-terminal region is ~300 residues and highly acidic. The Spt6 core contains five recognizable domains and is structurally similar to the bacterial Tex protein (Johnson et al., 2008). The C-terminal region of Spt6 contains a tandem SH2 domain comprising the only identified SH2 domains in yeast (Close et al., 2011; Diebold et al., 2010b; Sun et al., 2010). The N-terminal region of Spt6 is critical for Spt6 function and is necessary for binding nucleosomes (McDonald et al., 2010). Interestingly, a 30 residue segment in this region also binds the transcription factor Spn1 in a manner that is competitive with nucleosomes, suggesting a potential regulatory switch (McDonald et al., 2010). Consistent with this, mutating the Spt6-Spn1 interface causes defects in maintaining repressive chromatin that can be suppressed by H2A-H2B mutations as well as overexpression of Spn1 (McCullough et al., 2015). The Spt6 interaction with nucleosomes requires the presence of the small HMGB family member Nhp6 that

is thought to bind DNA and weaken the histone-DNA contacts (Stillman, 2010). This suggests that the histone chaperone activity of Spt6 may involve competing with DNA for histone binding as has been observed for other chaperones (Andrews et al., 2010; Kemble et al., 2015; Winkler et al., 2011). In order to gain insight into the molecular details of the histone chaperone activity of Spt6, and to assess the potential regulatory role of Spn1 in this process, we utilized biochemical binding and competition assays. The data presented here suggest that Spt6 has multiple $(H3\text{-}H4)_2$ binding sites, one of which resides in the Spn1 binding region. The Spt6-$(H3\text{-}H4)_2$ interaction is competitive with both Spn1 and DNA. Mutations that specifically disrupt the interaction with $(H3\text{-}H4)_2$ cause synthetics defects in combination with mutations in Spn1 and other histones chaperones *in vivo*, confirming the biological importance of the interaction. Overall, these data support the model that Spn1 regulates the histone chaperone activity of Spt6, and demonstrate that Spt6 utilizes a DNA-competition mechanism for chaperoning histones.

## Results

### Mapping the Spt6-$(H3\text{-}H4)_2$ Binding Interface

Spt6 has been reported to bind H3-H4 (Bortvin and Winston, 1996; McCullough et al., 2015). In order to investigate this interaction more quantitatively, we used fluorescence anisotropy to measure binding affinities of different Spt6 constructs to H3-H4 derived from *Xenopis lavis*. For larger Spt6 constructs, Spt6 was titrated into fluorescently labeled histones (Figure 5.1A). For $Spt6^{239\text{-}268}$, histones were titrated into fluorescently labeled Spt6 (Figure

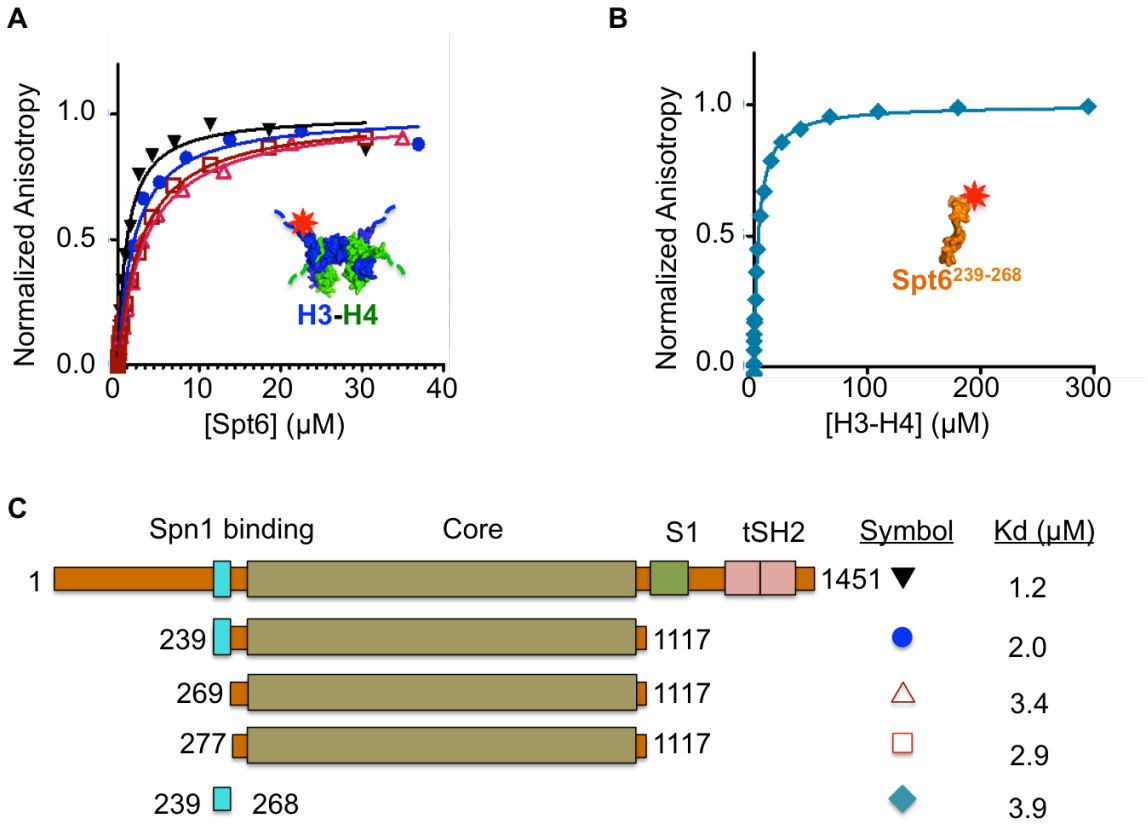



**Figure 5.1. The Spt6-Spn1 binding region is sufficient for binding histones H3-H4.**

A) Fluorescence anisotropy binding assay with different Spt6 constructs titrated against fluorescently labeled H3-H4.

B). Fluorescence anisotropy binding assay with H3-H4 titrated against fluorescently labeled $Spt6^{239-268}$.

C) Schematic showing the Spt6 constructs tested for binding and the respective affinities.

5.1B). All constructs containing the Spn1 binding region (residues 239-268) bound $(H3\text{-}H4)_2$ with an equilibrium dissociation constant ($K_D$) between 1 and 4 μM, indicating that these residues are sufficient for binding (Figure 5.1C). Surprisingly, removing these residues did not affect binding, suggesting that an additional independent binding site exists. We focused on the binding site in the Spn1 binding region for this work.

Characterizing the H3-H4 Interaction with Spt6

Because histones H3-H4 form a symmetric tetramer, there is a possibility for two identical binding sites that would allow two Spt6 molecules to bind one tetramer. Using isothermal titration calorimetry (ITC), the stoichiometry of $Spt6^{239\text{-}489}$ binding to $(H3\text{-}H4)_2$ was found to be 1:1 Spt6:$(H3\text{-}H4)_2$ (Figure 5.2A). The $K_D$ in this assay was ~2 μM, consistent with our anisotropy assays. There are two possible explanations for the observed stoichiometry: 1) the $(H3\text{-}H4)_2$-Spt6 binding site spans adjacent dimers or 2) two equivalent binding sites exist, but binding of one Spt6 molecule occludes the other binding site. In order to test these models, we used fluorescence anisotropy to measure the affinity of Spt6 for dimeric H3-H4 that are unable to form tetramers because of a mutation in the H3-H3' interface (Figure 5.2B and C). H3-H4 dimers bound Spt6 with an affinity similar to tetramers, indicating that the tetramer interface is not required for interaction with Spt6. These results support a model where binding of one Spt6 to a H3-H4 tetramer occludes an equivalent binding site on the adjacent dimer.

Histones have disordered N-terminal tails that are subject to posttranslational modifications and recruit accessory proteins. In order to

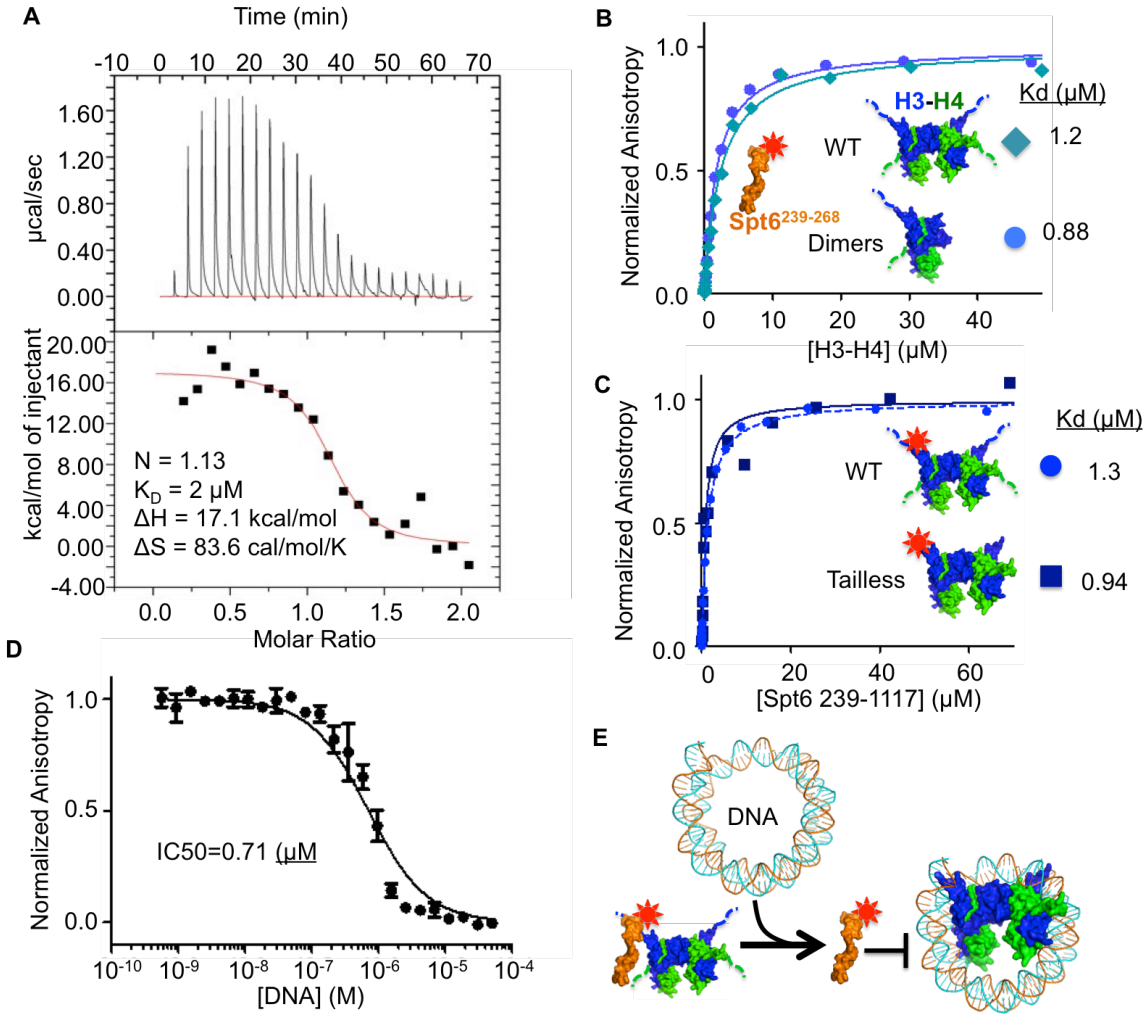



**Figure 5.2. Spt6 binds globular H3-H4 with a 1:1 stoichiometry and is competitive with DNA.** Seth McDonald performed the experiments shown in this figure. The slightly tighter affinities observed in this figure can be explained by person-to-person variation.

A) ITC with Spt6$^{239\text{-}489}$ and $(H3\text{-}H4)_2$. Raw data are shown in the top panel and the binding isotherm in the bottom panel. Thermodynamic parameters and stoichiometry are shown.

B) Fluorescence anisotropy assay. Wild type $(H3\text{-}H4)_2$ or dimeric H3$^{H114D}$-H4 were titrated against fluorescently labeled Spt6$^{239\text{-}268}$.

C) Fluorescence anisotropy with Spt6$^{239\text{-}1117}$ titrated against either wild type $(H3\text{-}H4)_2$ or tailless $(H3^{27\text{-}135}\text{-}H4^{20\text{-}102})_2$.

D) Competition fluorescence anisotropy assay. Purified DNA was titrated against constant concentrations of Spt6$^{239\text{-}268}$ (~1 nM) and $(H3\text{-}H4)_2$ (5 µM).

E) Schematic illustrating the competition fluorescence anisotropy assay described in D.

determine if the H3-H4 tails are necessary for binding with Spt6, we tested binding to H3-H4 lacking the N-terminal 26 and 19 residues respectively. The "tailless" histones bound labeled $Spt6^{239-268}$ as well as full-length $(H3\text{-}H4)_2$ (Figure 5.2D), supporting an interaction with the globular region of H3-H4.

Histone chaperones Nap1 and FACT act by competing with DNA for histone binding (Andrews et al., 2010; Kemble et al., 2015; Winkler et al., 2011). The requirement of Nhp6 for Spt6 binding to nucleosomes suggests that Spt6 uses a similar mechanism. We used a competition fluorescence anisotropy assay to directly determine if DNA competes with Spt6 for binding $(H3\text{-}H4)_2$. Labeled $Spt6^{239-268}$ and $(H3\text{-}H4)_2$ were kept at constant concentrations while the concentration of double stranded 147 bp "601" DNA was varied. DNA efficiently competed with Spt6 for $(H3\text{-}H4)_2$ binding with an IC50 of 0.71 μM (Figure 5.2E). This suggests a model where Spt6 promotes nucleosome assembly by blocking nonproductive interactions between $(H3\text{-}H4)_2$ and DNA.

Spt6 Interaction with H3-H4 is Competitive with Spn1

It has been demonstrated that Spn1 prevents Spt6 interaction with nucleosomes (McDonald et al., 2010). Because Spn1 and H3-H4 bind the same region on Spt6, it is likely that Spn1 competes with H3-H4 for binding. We used a competitive fluorescence anisotropy assay to test this hypothesis. Fluorescently labeled H3-H4 and $Spt6^{239-1117}$ were kept at constant concentrations while the concentration of Spn1 was varied. Consistent with our hypothesis, Spn1 competed with H3-H4 for Spt6 binding (Figure 5.3A). Unfortunately, a full competition curve could not be obtained because Spn1 associates with H3-H4 at

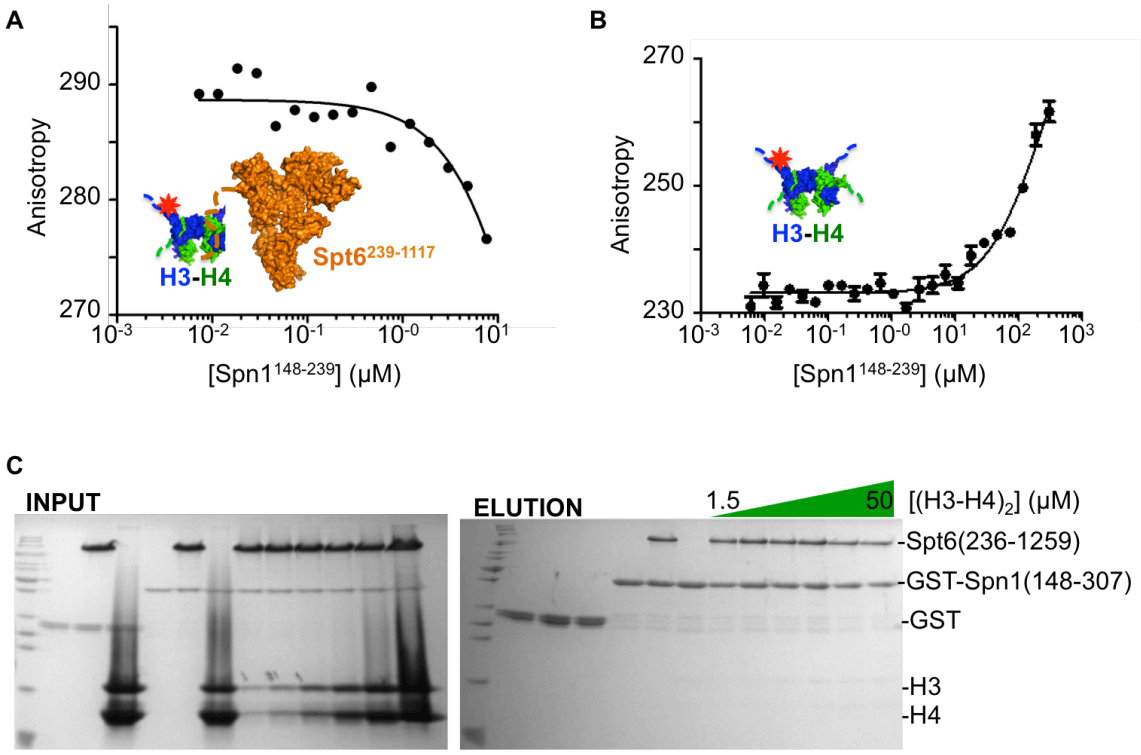



**Figure 5.3. Spn1 competes with $(H3\text{-}H4)_2$ for binding to Spt6.**
A) Competition fluorescence anisotropy assay. $Spn1^{148\text{-}239}$ was titrated against constant concentrations of fluorescently labeled H3-H4 (1 nM) and $Spt6^{239\text{-}1117}$ (5 μM).
B) Spn1 has weak affinity for H3-H4 in a fluorescence anisotropy binding assay. $Spn1^{148\text{-}239}$ was titrated against fluorescently labeled H3-H4.
C) GST pull-down assays with GST-$Spn1^{148\text{-}307}$, $Spt6^{236\text{-}1259}$, and $(H3\text{-}H4)_2$. H3-H4 does not pull-down in a ternary complex with Spn1 and Spt6, even though H3-H4 binds Spt6 under the same conditions (data not shown). Seth McDonald performed this experiment.

concentrations above 10 μM (Figure 5.3B). Competition between Spn1 and H3-H4 is further supported by the absence of a ternary complex in GST pull-downs under conditions that Spt6 is capable of interacting with both Spn1 and H3-H4 (Figure 3C). These data support a model where Spn1 regulates the histone chaperone activity of Spt6.

Mutations Disrupt the Spt6-$(H3\text{-}H4)_2$ Interaction *In Vitro*

In order to demonstrate specificity and develop tools to test our model *in vivo*, we identified Spt6 mutations that disrupt binding to H3-H4 *in vitro*. Scanning mutagenesis of the Spt6-Spn1 binding region was used to identify mutations that disrupt binding in a competition fluorescence anisotropy assay. Purified recombinant $Spt6^{239\text{-}268}$ proteins were titrated into constant concentrations of $H3^{H114D}$-H4 and fluorescently labeled $Spt6^{239\text{-}268}$. The $Spt6^{MYD245KKR}$ triple mutant reduced binding ~60 fold compared to wild type, demonstrating that the interaction is specific (Figure 5.4A). A single mutation of $Spt6^{Y246K}$ reduced binding ~6 fold compared to wild type. Importantly, this mutation did not affect binding to Spn1 (Figure 5.4B). We also tested the $Spt6^{F249K}$ mutation that is known to disrupt the interaction with Spn1 for binding to $(H3\text{-}H4)_2$. This mutation had no effect on binding (Figure 5.4C). The ability to specifically disrupt the interactions with $(H3\text{-}H4)_2$ and Spn1 independently allows for “separation of function” alleles to test the importance of each interaction *in vivo*.

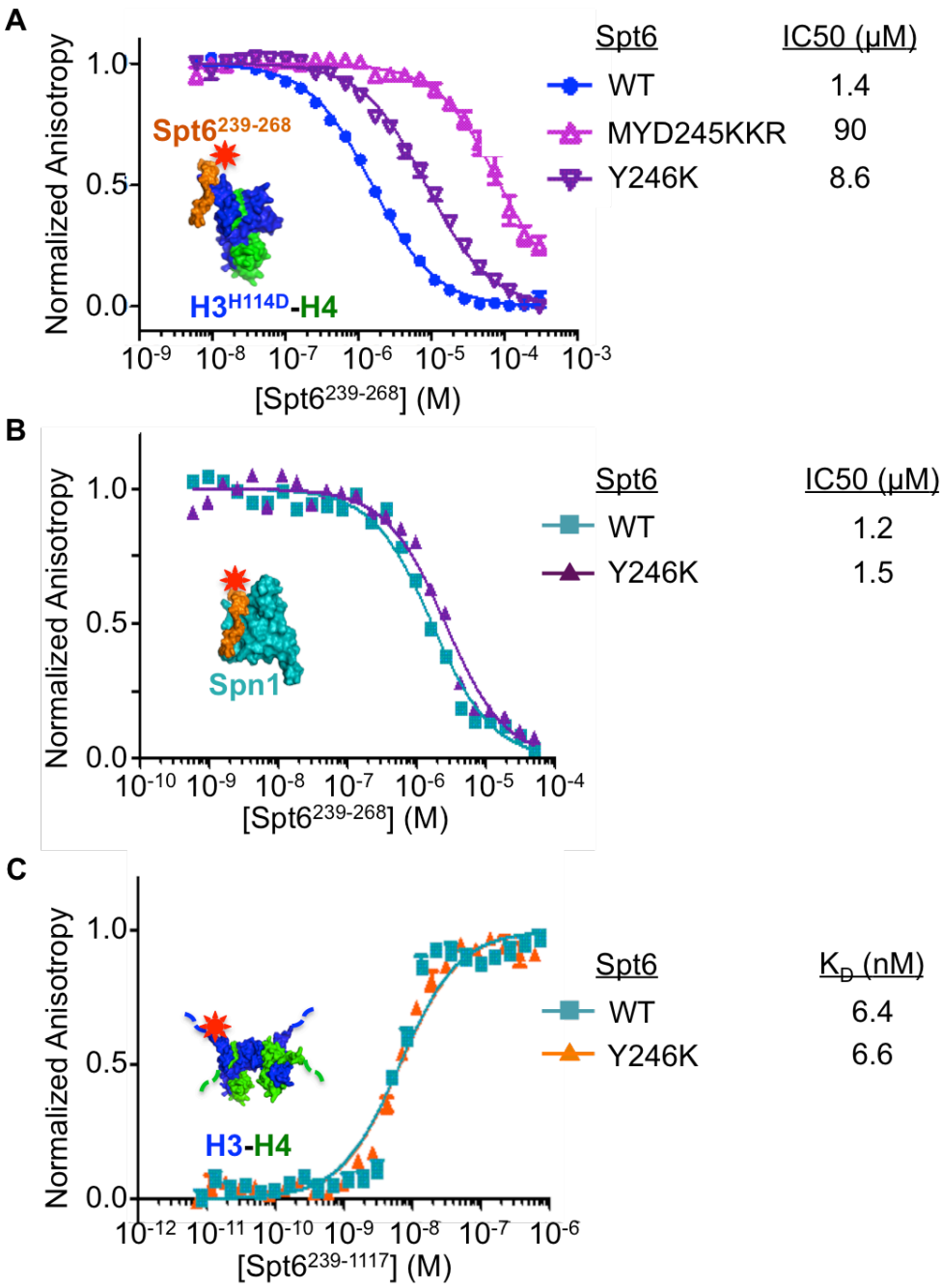



**Figure 5.4. Mutations specifically disrupt the Spt6-(H3-H4)$_2$ interaction *in vitro*.**

A) Competition fluorescence anisotropy assay. Spt6$^{239-268}$ mutants were titrated against constant concentrations of fluorescently labeled Spt6$^{239-268}$ and H3$^{H114D}$-H4. IC50's are reported.

B) Same as in A except Spt6 mutants were titrated against Spt6$^{239-268}$ and Spn1$^{148-239}$. IC50's are reported.

C) Fluorescence anisotropy assay in 200 mM NaCl. Spt6$^{239-1117}$ wild type or F249K were titrated against fluorescently labeled H3-H4. $K_D$'s are reported. Seth McDonald performed this experiment.

The Spt6-$(H3\text{-}H4)_2$ Interaction Is Important *In Vivo*

To confirm that the Spt6 interaction with H3-H4 observed *in vitro* is physiologically important, we introduced the *spt6-Y246K* mutation into the genome of yeast such that the mutant protein was expressed from its native promoter. *spt6-Y246K* did not cause a phenotype on its own under the conditions tested (Figure 5.5A). Combining *spt6-Y246K* with *pob3-Q308K*, a mutation in the Pob3 subunit of the histone chaperone FACT, results in impaired growth at elevated temperatures and sensitivity to the replication toxin hydroxyurea (HU). Similar phenotypes are observed when *spt6-Y246K* is crossed with a deletion of the histone chaperone Hpc2 or *spn1-F267E*, a mutation that disrupts the interaction with Spt6. Collectively, these results indicate that the Spt6-$(H3\text{-}H4)_2$ interface identified *in vitro* is important *in vivo*, and suggest a role in chaperoning histones.

## Discussion

Spt6 was previously shown to bind histones H3-H4 and promote nucleosome deposition *in vitro* (Bortvin and Winston, 1996; Formosa, 2012). In order to better understand the mechanism of Spt6-mediated nucleosome assembly, we have begun characterizing the histone binding sites on Spt6. We used fluorescence anisotropy to demonstrate that the same 30 residue region of Spt6 that binds Spn1 also binds histones H3-H4. These interaction are mutually exclusive, but can be selectively disrupted by point mutations. Disrupting the Spt6-$(H3\text{-}H4)_2$ interface *in vivo* causes growth defects when other modulators of chromatin structure are impaired, highlighting the importance of the interaction.

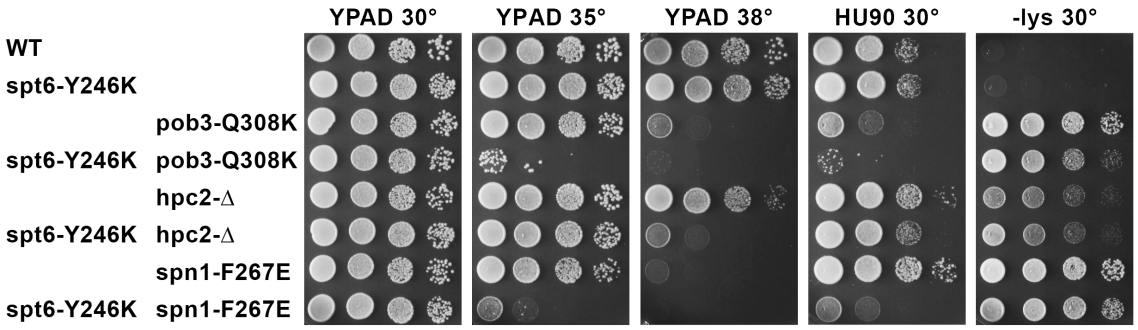



**Figure 5.5. The Spt6-$(H3\text{-}H4)_2$ interaction is important *in vivo*.**
Spot plates of yeast harboring the *spt6-Y246K* mutation. These experiments were performed in the laboratory of Tim Formosa.

$(H3\text{-}H4)_2$ is a symmetric dimer of H3-H4, suggesting the possibility of two identical binding sites. However, Spt6 binds H3-H4 dimers and tetramers with equivalent affinity and tetramers with a 1:1 stoichiometry. These results suggest that two binding sites should exist, but binding of Spt6 to one H3-H4 site occludes the second binding site. This would require that the binding site be close to the tetramer interface. Binding is also competitive with DNA, consistent with the requirement of Nhp6 for Spt6 binding to nucleosomes (McDonald et al., 2010). The most likely H3-H4 binding site that would allow for competition with DNA and be near the tetramer interface is adjacent to the nucleosome dyad. In support of this, the docking program ClusPro 2.0 (Comeau et al., 2004a, b) predicts a binding site in this region (data not shown).

Our mapping reveals that the Spt6-Spn1 binding region is sufficient but not necessary for binding H3-H4, suggesting the existence of an additional binding site. This binding site appears to be independent from the Spn1 binding region as there is little difference in affinity when the Spn1 binding region is removed. If the binding sites were coordinated, we would expect to see significantly tighter binding when both sites are present due to avidity effects. Consistent with the existence of other binding sites, a mutation that disrupts the $Spt6^{239\text{-}268}$-$(H3\text{-}H4)_2$ interaction does not have a phenotype unless combined with mutations in other proteins that modulate chromatin structure. On the other hand, mutations that impair the Spt6-Spn1 interaction cause severe growth phenotypes (McDonald et al., 2010).

Collectively, the work presented in this chapter expands our

understanding of how Spt6 chaperones histones. These results suggest a model where Spt6 promotes nucleosome assembly by preventing non-nucleosomal histone-DNA contacts. This appears to be accomplished through multiple independent binding sites, one of which overlaps with the Spn1 binding region. Additional biochemical studies are necessary to map the other binding site(s) and determine its contribution to nucleosome assembly. Spn1 may be actively engaged in chaperoning histones, possibly by disengaging Spt6 from nucleosomes in order to recycle it for subsequent rounds of reassembly. Alternatively, Spn1 may regulate Spt6 recruitment to sites of nucleosome reassembly. Regardless, Spt6 interactions with Spn1 and histones are important *in vivo* and additional studies are required to understand the mechanistic basis of nucleosome reassembly by Spt6.

## Materials and Methods

### Protein Expression and Purification

Spt6 and Spn1 proteins were expressed and purified as described previously (McDonald et al., 2010) with one modification: Spt6 constructs with C-terminal truncations at or before residue 489 were purified via anion-exchange chromatography (5 mL Q HP, GE Healthcare) in place of a Heparin cation-exchange column.

*Xenopus laevis* histones were expressed recombinantly and purified from inclusion-bodies as described elsewhere (Luger et al., 1999). ‘Tailless’ histones were prepared by inserting a PreScission Protease site such that upon protease treatment (after refolding), a GP dipeptide remains N-terminally fused to the

following residues: $H3^{27-135}$ and $H4^{20-102}$. Mutating H3 H114, a residue in the H3-H3 tetramerization interface, to an aspartate produced dimeric histones (Andrews et al., 2008). Fluorescent labeling of histone $H4^{T71C}$ and $Spt6^{239-268}$ with Oregon Green-488 maleimide (Invitrogen) was performed according to manufacturers instructions. $Spt6^{239-268}$ was labeled on a cysteine residue introduced by site-directed mutagenesis in the linker sequence that remains after TEV protease treatment.

DNA Preparation

The 147 bp “601” nucleosome positioning sequence fragment was purified from large-scale DNA plasmid preps of pST55 (a gift from Song Tan), followed by digestion with EcoRV, and gel purification.

Fluorescence Anisotropy

All fluorescence anisotropy assays were incubated for 1 hour and measured at 25°C in 384-well flat bottom black (nonbinding surface) microplates (Corning). Purified Spt6 proteins were titrated in 1.6-1.8 fold serial dilutions against a constant concentration (0.5-1 nM) of labeled histones in anisotropy buffer (20 mM Tris pH 7.5, 750 mM NaCl, 5% glycerol, and 1 mM DTT). For Spt6 F249K, assays were performed in low salt anisotropy buffer with 200 mM NaCl. $K_D$ values were determined by fitting the data in GraphPad Prism® using nonlinear least squared regression against $A = A_T / (2 \times [pep]) \times (([pro] + [pep] + K_D) - (([pro] + [pep] + K_D)^2 - 4 \times [pro] \times [pep])^{1/2})$ (LiCata and Wowor, 2008), where A is the measured anisotropy, $A_T$ is the total change in anisotropy, [pep] is

the peptide concentration, and [pro] is the protein concentration. DNA competition assays were performed in anisotropy buffer by titration of DNA against a constant concentration of labeled $Spt6^{238-269}$ (1 nM) and H3-H4 (5 μM). Competition assays with Spt6 mutants were performed in anisotropy buffer by titration of Spt6 proteins against constant concentrations of labeled $Spt6^{239-268}$ (1 nM) and $H3^{H114D}$-H4 (5 μM). IC50 values were determined by fitting to the standard IC50 equation in GraphPad Prism®.

Isothermal Titration Calorimetry

Purified recombinant proteins were dialyzed at 4°C overnight against 2 L of degassed ITC buffer (20 mM HEPES pH 7.5, 750 mM NaCl, 0.5 mM EDTA, 5 % glycerol, and 1 mM DTT). Titrations were done at 25°C on an iTC200 (Microcal) and spaced 180 s apart. $Spt6^{239-489}$ was tested with 1.8 μL injections of 1.25 mM Spt6 into 0.125 mM $(H3\text{-}H4)_2$. Origin software (Microcal) was used for data analysis and determination of thermodynamic parameters.

Yeast Strains

Yeast strains were constructed and phenotyped in the laboratory of Tim Formosa.

## References

Andrews, A.J., Chen, X., Zevin, A., Stargell, L.A., and Luger, K. (2010). The histone chaperone Nap1 promotes nucleosome assembly by eliminating nonnucleosomal histone DNA interactions. Mol Cell *37*, 834-842.

Andrews, A.J., Downing, G., Brown, K., Park, Y.J., and Luger, K. (2008). A thermodynamic model for Nap1-histone interactions. J Biol Chem *283*, 32412-32418.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J *28*, 1067-1077.

Bortvin, A., and Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. Science *272*, 1473-1476.

Bucheli, M.E., and Buratowski, S. (2005). Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. EMBO J *24*, 2150-2160.

Clark-Adams, C.D., and Winston, F. (1987). The SPT6 gene is essential for growth and is required for delta-mediated transcription in Saccharomyces cerevisiae. Mol Cell Biol *7*, 679-686.

Close, D., Johnson, S.J., Sdano, M.A., McDonald, S.M., Robinson, H., Formosa, T., and Hill, C.P. (2011). Crystal structures of the S. cerevisiae Spt6 core and C-terminal tandem SH2 domain. J Mol Biol *408*, 697-713.

Comeau, S.R., Gatchell, D.W., Vajda, S., and Camacho, C.J. (2004a). ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res *32*, W96-99.

Comeau, S.R., Gatchell, D.W., Vajda, S., and Camacho, C.J. (2004b). ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics *20*, 45-50.

Compagnone-Post, P.A., and Osley, M.A. (1996). Mutations in the SPT4, SPT5, and SPT6 genes alter transcription of a subset of histone genes in Saccharomyces cerevisiae. Genetics *143*, 1543-1554.

DeGennaro, C.M., Alver, B.H., Marguerat, S., Stepanova, E., Davis, C.P., Bahler, J., Park, P.J., and Winston, F. (2013). Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. Mol Cell Biol *33*, 4779-4792.

Diebold, M.L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., and Romier, C. (2010). Noncanonical tandem SH2 enables interaction of elongation factor Spt6 with RNA polymerase II. J Biol Chem *285*, 38389-38398.

Dronamraju, R., and Strahl, B.D. (2014). A feed forward circuit comprising Spt6, Ctk1 and PAF regulates Pol II CTD phosphorylation and transcription elongation. Nucleic Acids Res *42*, 870-881.

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A.*, et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol Cell Biol *24*, 3324-3336.

Formosa, T. (2012). The role of FACT in making and breaking nucleosomes. Biochim Biophys Acta *1819*, 247-255.

Gurard-Levin, Z.A., Quivy, J.P., and Almouzni, G. (2014). Histone chaperones: assisting histone traffic and nucleosome dynamics. Annu Rev Biochem *83*, 487-517.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev *12*, 357-369.

Ivanovska, I., Jacques, P.E., Rando, O.J., Robert, F., and Winston, F. (2011). Control of chromatin structure by spt6: different consequences in coding and regulatory regions. Mol Cell Biol *31*, 531-541.

Johnson, S.J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S.L., and Hill, C.P. (2008). Crystal structure and RNA binding of the Tex protein from Pseudomonas aeruginosa. J Mol Biol *377*, 1460-1473.

Kaplan, C.D., Holland, M.J., and Winston, F. (2005). Interaction between transcription elongation factors and mRNA 3'-end formation at the Saccharomyces cerevisiae GAL10-GAL7 locus. J Biol Chem *280*, 913-922.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.

Kemble, D.J., McCullough, L.L., Whitby, F.G., Formosa, T., and Hill, C.P. (2015). FACT disrupts nucleosome structure by binding H2A-H2B with conserved peptide motifs. Mol Cell *60*, 294-306.

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Curr Opin Genet Dev *21*, 175-186.

LiCata, V.J., and Wowor, A.J. (2008). Applications of fluorescence anisotropy to the study of protein-DNA interactions. Methods Cell Biol *84*, 243-262.

Lorch, Y., LaPointe, J.W., and Kornberg, R.D. (1992). Initiation on chromatin templates in a yeast RNA polymerase II transcription system. Genes Dev *6*, 2282-2287.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol *119*, 1-16.

McCullough, L., Connell, Z., Petersen, C., and Formosa, T. (2015). The abundant histone chaperones Spt6 and FACT collaborate to assemble, inspect, and maintain chromatin structure in Saccharomyces cerevisiae. Genetics *201*, 1031-1045.

McDonald, S.M., Close, D., Xin, H., Formosa, T., and Hill, C.P. (2010). Structure and biological importance of the Spn1-Spt6 interaction, and its regulatory role in nucleosome binding. Mol Cell *40*, 725-735.

Park, Y.J., and Luger, K. (2008). Histone chaperones in nucleosome eviction and histone exchange. Curr Opin Struct Biol *18*, 282-289.

Perales, R., Erickson, B., Zhang, L., Kim, H., Valiquett, E., and Bentley, D. (2013). Gene promoters dictate histone occupancy within genes. EMBO J *32*, 2645-2656.

Ransom, M., Dennehey, B.K., and Tyler, J.K. (2010). Chaperoning histones during DNA replication and repair. Cell *140*, 183-195.

Stillman, D.J. (2010). Nhp6: a small but powerful effector of chromatin structure in Saccharomyces cerevisiae. Biochim Biophys Acta *1799*, 175-180.

Sun, M., Lariviere, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). J Biol Chem *285*, 41597-41603.

Winkler, D.D., Muthurajan, U.M., Hieb, A.R., and Luger, K. (2011). Histone chaperone FACT coordinates nucleosome interaction through multiple synergistic binding events. J Biol Chem *286*, 41883-41892.

Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes Dev *21*, 160-174.

Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. Genes Dev *22*, 3422-3434.

# CHAPTER 6

# CONCLUSIONS AND ONGOING RESEARCH

## Summary

Spt6 is conserved throughout eukaryotes, is homologous to the prokaryotic transcription factor Tex, and is essential in *S. cerevisiae*. Spt6 influences several facets of gene expression including transcription, mRNA processing and export, posttranslational modification of histones, and nucleosome occupancy. The mechanistic details of Spt6 involvement in these processes are unknown, but interactions with proteins and nucleic acid are likely important. The focus of this dissertation has been to identify and characterize Spt6 binding partners. Chapter 2 reported a crystal structure of Spt6 including a C-terminal tandem SH2 domain. The tSH2 structure and accompanying biochemistry formed the basis for the work in Chapters 3 and 4.

In Chapter 3, I showed that, contrary to popular belief, Spt6 tSH2 binds the phosphorylated Rpb1 linker but not the CTD. The interaction is phosphorylation dependent and biologically important for maintaining repressive chromatin. The work in Chapter 4 identified phosphorylated Tom1 as a previously unidentified Spt6 tSH2 binding partner that may connect Spt6 to mRNA export. In addition, an 8 Å cryo-EM reconstruction revealed that ~300 kDa of the 375 kDa Tom1 protein are substantially folded, forming the basis for a mechanistic

understanding of Tom1.

In Chapter 5, in collaboration with Seth McDonald, I characterized the interaction between histones H3-H4 and the Spt6-Spn1 binding region. Those studies support a model in which Spt6 competes with DNA for H3-H4 binding to prevent non-nucleosomal DNA-histone interactions. This chapter describes ongoing studies of Spt6 and Tom1 and discusses additional experiments designed to further understand the diverse roles of these multifaceted proteins.

## Function of the Spt6 Core

The core of Spt6 is structurally similar to the prokaryotic transcription factor Tex (Johnson et al., 2008) and contains five domains with structural homology to protein-binding and nucleic acid-binding domains (Close et al., 2011). Therefore, it is likely that the Spt6 core participates in direct binding interactions with other proteins and/or nucleic acids. Consistent with this, Spt6 binds DNA *in vitro*, although it is unclear if this is biologically relevant (Close et al., 2011). At this point, no proteins have been identified that bind specifically with the Spt6 core, although I believe that interacting proteins are likely to be fundamental to Spt6 function. Therefore, their identification will be a key step in advancing understanding of Spt6 functions and mechanisms. In order to identify Spt6 core binding partners, a proteomics approach similar to that utilized for the tSH2 domain (Chapter 4) should be pursued. The Spt6 core can be overexpressed in yeast with tandem affinity tags and purified under physiological conditions to identify co-purifying species. Identifying and characterizing these binding interactions is expected to shed light on how Spt6 participates in such a

wide variety of biological processes.

## Spt6 tSH2 Phosphotyrosine Binding

The work presented in Chapter 3 identifies a biologically relevant interaction between Spt6 tSH2 and the Rpb1 linker. Preliminary evidence suggests that an additional phosphorylation dependent interaction with the Rpb1 linker is mediated by the putative phosphotyrosine binding pocket. In our crystal structure, the N-terminal residues of the linker are approaching the highly conserved phosphotyrosine binding pocket of the canonical nSH2 domain. Modeling additional Rpb1 residues into our crystal structure reveal that a conserved tyrosine (Rpb1 Y1473) would be in good position to sit in the phosphotyrosine binding pocket (Figure 6.1A). In support of this, phosphorylation of this tyrosine has been observed in high throughput mass spectrometry experiments (Albuquerque et al., 2008). Another residue (Rpb1 T1471) two amino acids N-terminal to Y1473 is also strictly conserved, has also been observed to be phosphorylated by mass spectrometry (Albuquerque et al., 2008), and is marginally important for Spt6 tSH2 binding in a far western blot (Chapter 3). In support of Rpb1 Y1473 and T1471 contributing to Spt6 binding, mutation of either residue in yeast has strong synthetic defects with *rpb1-S1493A* and Spt6 mutations in the phosphoserine binding pocket (data not shown). The same synthetic defect is seen when a mutation in the putative Spt6 phosphotryosine binding pocket (*spt6-R1282H*) is combined with mutations that impair phosphoserine binding. The growth defect observed when both Spt6 tSH2 phosphate binding pockets are mutated is similar to the defect caused by

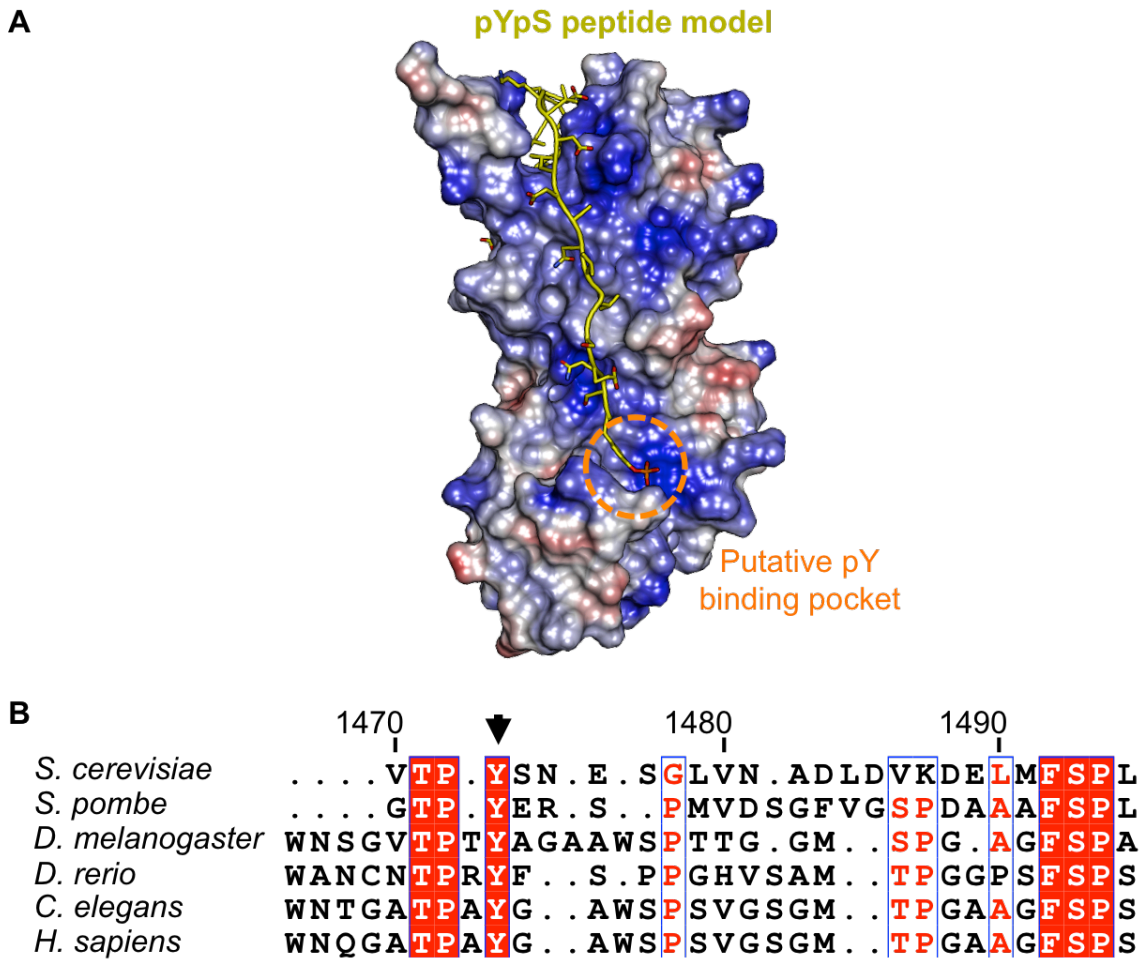



**Figure 6.1. Phosphorylation of Rpb1 Y1473 may contribute to Spt6 tSH2 binding through the canonical phosphotyrosine binding pocket.**
A) Model of Rpb1 pY1473 docked into the Spt6 tSH2 phosphotyrosine binding pocket. Spt6 tSH2 is colored according to electrostatic surface potential. The putative phosphotyrosine binding pocket is highlighted with an orange dashed circle. Rpb1 is shown as yellow sticks. The Spt6 tSH2-Rpb1 linker structure was modified by modeling additional residues at the linker N-terminus. Modeling was performed in COOT (Emsley et al., 2010) and the figure was generated using UCSF Chimera (Pettersen et al., 2004).
B) Alignment of a portion of the Rpb1 linker across eukaryotes. The presence of a tyrosine (*S. cerevisiae* 1473, black arrowhead) 18-20 residues N-terminal to S1493 is strictly conserved. Sequences were aligned using T-Coffee (Notredame et al., 2000) and displayed using ESPript (Robert and Gouet, 2014). White characters on a red background represent invariant residues and red font on a white background represents highly similar (global similarity score $\geq$ 0.7) residues.

complete deletion of the entire tSH2 domain. An attractive model to explain these observations is that the Spt6 tSH2-Rpb1 linker interface extends beyond the surface characterized in Chapter 3 and likely involves additional phosphorylation events. Linker derived peptides with different combinations of phosphorylation marks are being synthesized to test for binding. Tightly binding peptides will be further characterized by x-ray crystallography. In addition, ChIP studies are underway to determine the impact of mutating the Spt6 phosphate binding pockets on Spt6 recruitment to genes.

## Spt6 Interaction with the RNAPII Core

Spt6 displays a relatively rare ability to stimulate RNAPII elongation rates *in vitro* on nucleosome free DNA (Endoh et al., 2004). It is difficult to imagine how Spt6 binding to the flexible Rpb1 linker could be responsible for this activity. It is possible that Spt6 tSH2 binding to the linker destabilizes association of Rpb4/7 with the core to enhance elongation. This seems unlikely, as it has never been demonstrated that the linker contacts with Rpb4/7 observed in the *S. pombe* crystal structure are functionally important. Another possibility is that Spt6 interacts directly with the RNAPII core in order to facilitate more efficient transcription elongation. Consistent with this, preliminary pull-down experiments suggest that Spt6 tSH2 binds the RNAPII core because substantial residual copurification is apparent when the Rpb1 CTD and linker are absent (Figure 6.2A and B). There are two possible explanations why this interaction was not observed in the far western blots described in Chapter 3. The first possibility is that the concentrations used in the far western blot are significantly lower than

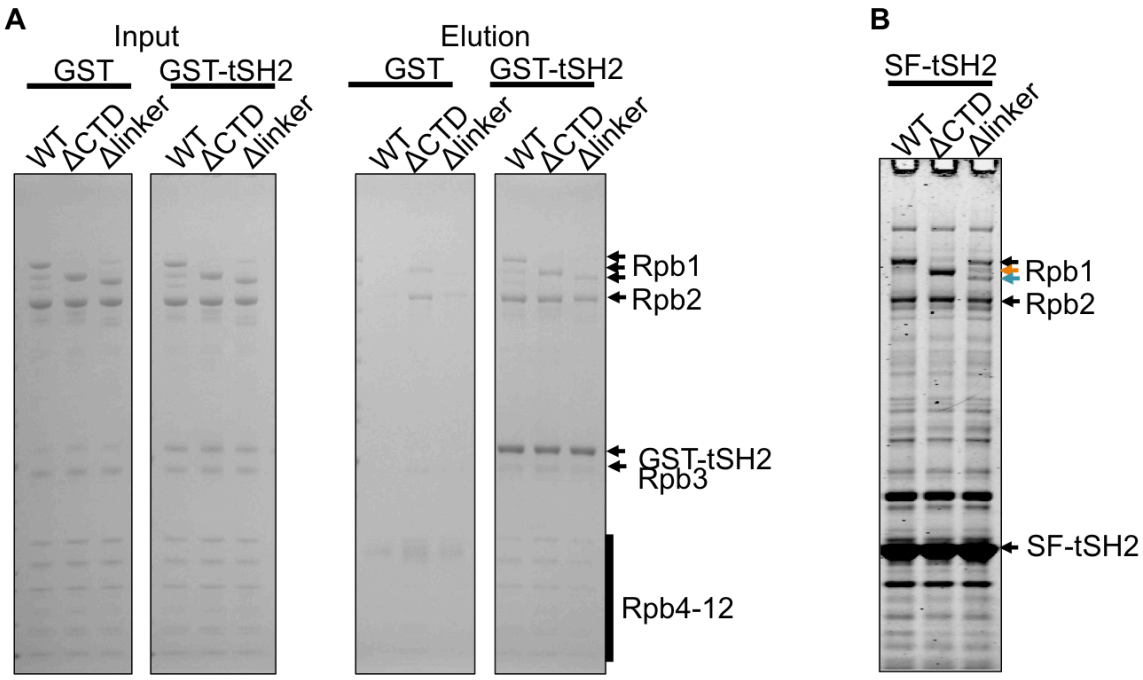



**Figure 6.2. Spt6 tSH2 interacts with the RNAPII core in pull-down experiments**.

A) GST pull-downs using purified proteins. Wild type and truncated versions of purified RNAPII were passed over GST-Spt6$^{1223\text{-}1451}$ immobilized on glutathione resin. Bound proteins were resolved by SDS-PAGE and stained with coomassie. B) Tandem affinity purification of strep-FLAG tagged Spt6$^{1223\text{-}1451}$ (SF-tSH2) that was overexpressed in yeast strains with PreScission Protease sites inserted at different locations in Rpb1. The Rpb1 CTD and linker-CTD were cleaved off (ΔCTD, orange arrow; and Δlinker, cyan arrow) by incubation with PreScission Protease during the first purification step. Co-purifying proteins were resolved by SDS-PAGE and stained with coomassie.

the pull-down experiment, so that weaker interactions would not be detected by far western blot. The second explanation is that the interaction requires a structured RNAPII interface that is not refolded in the far western blot. Regardless, pull-downs using folded proteins suggest that a binding site exists for Spt6 on the RNAPII core, and this may explain how Spt6 stimulates transcription elongation.

We have attempted to visualize this interaction using cryo-EM, but were unable to identify density corresponding to Spt6 tSH2. We were not limited by resolution, as the average resolution of the reconstruction was 4.2 Å. We observed significant structural heterogeneity of RNAPII in our reconstruction. In fact, over half of our particles separated into 2D class averages that appeared to contain only half of RNAPII. This suggests that exposure to the air-water interface in the moments prior to vitrification could be causing our particles to fall apart, even with crosslinking. Therefore, it is possible that we are unable to sufficiently resolve the heterogeneity to be able to identify a group of homogenous particles that contain Spt6 tSH2 bound to the core. Alternatively, Spt6 tSH2 could be interacting with a flexible loop on the RNAPII core, so it is not well resolved. An additional possibility is that the interaction observed in pull-downs is nonspecific. Even if the Spt6 tSH2 interaction with the core is nonspecific, other parts of Spt6 could be binding the RNAPII core. This is supported by the observation that deletion of Spt6 tSH2 domain diminishes, but does not eliminate Spt6 recruitment to genes (Burugula et al., 2014; Mayer et al., 2010). In addition, the core of Spt6 is homologous the prokaryotic factor Tex

which co-purifies with bacterial RNA Polymerase (Simon Dove, unpublished data). Spt6 also shares overlapping domains with the mitochondrial transcription factor TEFM that directly binds the mitochondrial RNA Polymerase (POLRMT) (Minczuk et al., 2011; Posse et al., 2015). Because the Spt6 core may contribute to RNAPII binding, future attempts at cryo-EM should use a larger version of Spt6 that encompasses the core and tSH2 domain. In addition, including a nucleic acid template is advisable to more closely mimic a transcribing RNAPII complex. Finally, using grids with a thin layer of carbon could reduce the number of half particles by pulling particles away from the air-water interface prior to vitrification.

## Biochemical and Structural Characterization of Tom1

The identification of Tom1 as an Spt6 tSH2 binding partner (Chapter 4) was unanticipated, but the interaction appears to be biologically important. While the specific process the interaction is important for is unclear, Tom1 does have several activities that overlap with Spt6 including mRNA export, excess histone degradation, and cell cycle regulation. Of these, mRNA export is the most attractive because of overlapping genetic and physical association with Yra1 (Burckin et al., 2005; Estruch et al., 2009; Iglesias et al., 2010; Utsugi et al., 1999; Yoh et al., 2007). The most obvious experiment to test for mRNA export defects is to use fluorescence *in situ* hybridization (FISH) to monitor mRNA localization when the Tom1-tSH2 binding region is deleted in yeast. Testing the importance of the interaction for histone degradation is challenging, because the pool of free histones is relatively small compared to total cellular histone levels

(0.6% H3) (McCullough et al., 2013). Our collaborator Tim Formosa has developed tools that will be useful for identifying changes in free histone levels when the Spt6-Tom1 interaction is impaired (McCullough et al., 2013; Xin et al., 2009). In addition, flow cytometry could be used to look for defects in cell cycle progression when the binding site is mutated. It may also be insightful to identify the pathway that regulates Tom1 phosphorylation, starting with the kinase. The sequence surrounding Tom1 S1943 resembles a casein kinase II (CKII) site (Glover, 1998). Since CKII has been associated with Spt6 function (Krogan et al., 2002), it should be the first candidate investigated.

In the event that a targeted approach does not reveal the biological importance of the Spt6-Tom1 interaction, a more global strategy will be required. Our collaborator, Mahesh Chandrasekharan, is currently initiating RNA-seq studies to identify genes that are regulated in a Tom1 dependent manner. Other avenues of exploration could include identification of Tom1 binding partners and substrates. Tom1 could be pulled out from yeast in physiologic buffer conditions followed by mass spectrometry to identify co-purifying species. In addition, identifying Tom1 substrates may offer insight into the pathways in which Tom1 participates. A recently described technique called ubiquitin-activated interaction traps (UBAITs) enables covalent attachment of E3 ligases to substrates by fusing ubiquitin to the C-terminus of the E3 (O'Connor et al., 2015). When the E3 ubiquitylates a substrate using the fused ubiquitin, an amide linkage is formed, enabling co-purification and subsequent identification by mass spectrometry. We are currently generating tools to use this method for Tom1.

## Spt6 Interactions with Histones

The work presented in Chapter 5 focuses on characterization of the interaction between the Spt6-Spn1 binding domain and histones $(H3\text{-}H4)_2$. Our binding data indicate that this is just one of multiple $(H3\text{-}H4)_2$ binding sites on Spt6. Identifying and characterizing the other binding site(s) is necessary to develop a complete mechanistic understanding of how Spt6 functions as a histone chaperone. ITC is likely the best technique for more thorough mapping, as it has the advantage of giving information about stoichiometry as well as binding affinity. This information will be useful in trying to distinguish between multiple independent sites with equivalent $K_D$. Once individual binding sites are mapped, structural studies can be pursued to determine where Spt6 binds histones and how these contacts may relate to nucleosome assembly.

The strategy outlined above should also be used to map Spt6 interactions with histones H2A-H2B. Spt6 is classically thought of as an H3-H4 chaperone (Bortvin and Winston, 1996), but it is also capable of binding H2A-H2B (McCullough et al., 2015). Preliminary ITC data from David Kemble suggests the presence of four H2A-H2B binding sites on Spt6, and data from Seth McDonald indicate that the Spn1 binding region may be one of them (data not shown). It is unclear why Spt6 has binding sites for more histones than are present in the nucleosome. Each binding site may be functionally important, or they may be artifacts of *in vitro* biochemistry performed with proteins that are prone to non-specific interactions. Careful biochemical characterization of the individual binding sites is necessary to determine if these interactions are specific.

Subsequently, mutations that disrupt each interface can be integrated into yeast to identify functional importance.

## Conclusions

The importance of Spt6 for a diverse set of biological processes makes it an exciting target for mechanistic studies. The work in this dissertation expands on the foundational work of several former members of the Hill lab: Sean Johnson, Devin Close, and Seth McDonald. The primary focus has been identification and characterization of Spt6 interactions with protein binding partners. Understanding the interactions that Spt6 participates in is fundamental to understanding how Spt6 is recruited to its different functional activities. The finding that Spt6 tSH2 binds RNAPII through the flexible linker rather than the CTD presents an important correction to the literature. The interaction is enhanced by previously unidentified phosphorylation of the linker and is important for maintaining repressive chromatin. We have also made the unexpected discovery that Spt6 tSH2 not only binds RNAPII but also Tom1. Determining the biological role of the interaction will advance understanding of Spt6 and Tom1 functions. Towards understanding Tom1 at a mechanistic level, we have solved a cryo-EM structure that reveals a substantially folded protein composed of helical repeats. Finally, biochemical characterization of the interaction between the Spt6-Spn1 binding region and histones H3-H4 reveals a role for competition with DNA in Spt6 mediated nucleosome assembly. Ongoing work includes structural, biochemical, and genetic studies to more thoroughly characterize Spt6 interactions with RNAPII, Tom1, and histones. Dissecting

these interactions will bring us closer to complete mechanistic understanding of how Spt6 functions in many facets of gene expression.

## References

Albuquerque, C.P., Smolka, M.B., Payne, S.H., Bafna, V., Eng, J., and Zhou, H. (2008). A multidimensional chromatography technology for in-depth phosphoproteome analysis. Mol Cell Proteomics *7*, 1389-1396.

Bortvin, A., and Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. Science *272*, 1473-1476.

Burckin, T., Nagel, R., Mandel-Gutfreund, Y., Shiue, L., Clark, T.A., Chong, J.L., Chang, T.H., Squazzo, S., Hartzog, G., and Ares, M., Jr. (2005). Exploring functional relationships between components of the gene expression machinery. Nat Struct Mol Biol *12*, 175-182.

Burugula, B.B., Jeronimo, C., Pathak, R., Jones, J.W., Robert, F., and Govind, C.K. (2014). Histone deacetylases and phosphorylated polymerase II C-terminal domain recruit Spt6 for cotranscriptional histone reassembly. Mol Cell Biol *34*, 4115-4129.

Close, D., Johnson, S.J., Sdano, M.A., McDonald, S.M., Robinson, H., Formosa, T., and Hill, C.P. (2011). Crystal structures of the S. cerevisiae Spt6 core and C-terminal tandem SH2 domain. J Mol Biol *408*, 697-713.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A.*, et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. Mol Cell Biol *24*, 3324-3336.

Estruch, F., Peiro-Chova, L., Gomez-Navarro, N., Durban, J., Hodge, C., Del Olmo, M., and Cole, C.N. (2009). A genetic screen in Saccharomyces cerevisiae identifies new genes that interact with mex67-5, a temperature-sensitive allele of the gene encoding the mRNA export receptor. Mol Genet Genomics *281*, 125-134.

Glover, C.V., 3rd (1998). On the physiological role of casein kinase II in Saccharomyces cerevisiae. Prog Nucleic Acid Res Mol Biol *59*, 95-133.

Iglesias, N., Tutucci, E., Gwizdek, C., Vinciguerra, P., Von Dach, E., Corbett, A.H., Dargemont, C., and Stutz, F. (2010). Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev *24*, 1927-1938.

Johnson, S.J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S.L., and Hill, C.P. (2008). Crystal structure and RNA binding of the Tex protein from Pseudomonas aeruginosa. J Mol Biol *377*, 1460-1473.

Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G., Kobor, M.S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J.F. (2002). RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. Mol Cell Biol *22*, 6979-6992.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol *17*, 1272-1278.

McCullough, L., Connell, Z., Petersen, C., and Formosa, T. (2015). The abundant histone chaperones Spt6 and FACT collaborate to assemble, inspect, and maintain chromatin structure in Saccharomyces cerevisiae. Genetics *201*, 1031-1045.

McCullough, L., Poe, B., Connell, Z., Xin, H., and Formosa, T. (2013). The FACT histone chaperone guides histone H4 into its nucleosomal conformation in Saccharomyces cerevisiae. Genetics *195*, 101-113.

Minczuk, M., He, J., Duch, A.M., Ettema, T.J., Chlebowski, A., Dzionek, K., Nijtmans, L.G., Huynen, M.A., and Holt, I.J. (2011). TEFM (c17orf42) is necessary for transcription of human mtDNA. Nucleic Acids Res *39*, 4284-4299.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217.

O'Connor, H.F., Lyon, N., Leung, J.W., Agarwal, P., Swaim, C.D., Miller, K.M., and Huibregtse, J.M. (2015). Ubiquitin-Activated Interaction Traps (UBAITs) identify E3 ligase binding partners. EMBO Rep *16*, 1699-1712.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Posse, V., Shahzad, S., Falkenberg, M., Hallberg, B.M., and Gustafsson, C.M. (2015). TEFM is a potent stimulator of mitochondrial transcription elongation in vitro. Nucleic Acids Res *43*, 2615-2624.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res *42*, W320-324.

Utsugi, T., Hirata, A., Sekiguchi, Y., Sasaki, T., Toh-e, A., and Kikuchi, Y. (1999). Yeast tom1 mutant exhibits pleiotropic defects in nuclear division, maintenance of nuclear structure and nucleocytoplasmic transport at high temperatures. Gene *234*, 285-295.

Xin, H., Takahata, S., Blanksma, M., McCullough, L., Stillman, D.J., and Formosa, T. (2009). yFACT induces global accessibility of nucleosomal DNA without H2A-H2B displacement. Mol Cell *35*, 365-376.

Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. Genes Dev *21*, 160-174.